



La poule qui chante

Lancement à l'international

Enjeu :

Première analyse des groupements de pays à cibler pour exporter nos poulets.

La poule qui chante souhaite étendre son marché à l'international.

Une première étude data-driven basée sur des indicateurs socio-politico-économiques nationaux permettra de définir des groupements de pays à fort potentiel pour lesquels une étude de marché plus approfondie serait pertinente par la suite.

Sommaire

- 1 **Analyse Pestel** pour déterminer les indicateurs à retenir
- 2 **Récupération des données** sur le site de la Food & Agriculture Organisation
- 3 **Analyse exploratoire** des données récupérées
- 4 1ère méthode de clustering : **classification ascendante hiérarchique** (ACH)
- 5 2ème méthode de clustering : **K-means**
- 6 **Comparaison** des résultats obtenus par les 2 méthodes
- 7 **Analyse en composantes principales** (PCA) : étude de la variabilité des pays et des liaisons entre les indicateurs
- 8 **Recommandations**

1

Analyse Pestel

Seuls les critères politiques, économiques, législatifs et sociaux ont été retenus pour cette première analyse.

Indicateurs retenus :

- **population et son taux d'évolution, taux d'urbanisation, PIB par habitant, distance**
- **stabilité politique,**
- **(concernant la viande de poulets) disponibilité, proportion consommée, indice des prix, index de croissance de la production, production et importation pour 1000 habitants, tarifs douaniers,**
- **régime alimentaire (part de protéines animales consommées),**
- **proportion de poulets (vivants) par rapport aux autres têtes de bétails**

L'analyse de marché ultérieure devra compléter cette étude préalable, notamment sur les critères écologiques et technologiques.



Récupération des données

Récupération
des données

Jointure
des données

Dataframe
final

Plusieurs sources ont été utilisées pour obtenir nos variables :

- <http://tariffdata.wto.org>, site de l'OMC pour les tarifs douaniers,
- <http://www.cepii.fr> pour le calcul des distances
- <https://www.fao.org/faostat/en/#data>, site de la FAO pour toutes les autres variables



	Area	alpha-3	region	dist	Population (1000)	pop_evo_rate	Urbanisation rate	gdp (\$) per capita	Food supply quantity (g/capita/day)	Import Quantity	Production	Chicken_meat_ratio	Livestock ratio (LSU %)	Price indices	production value	GPIIndex(2015)	Political stability	Animal protein ratio	AV Duties
0	Afghanistan	AFG	Asia	5590.38	36296.11	0.34	0.25	513.09	5.00	28937.00	27850.00	0.10	2.04	113.83	NaN	82.65	-2.75	0.19	32.10
1	Albania	ALB	Europe	1603.53	2884.17	-0.05	0.60	4514.20	29.00	17280.00	12912.00	0.14	10.54	105.13	26080000.00	157.14	0.38	0.54	10.00
2	Algeria	DZA	Africa	1340.39	41389.19	0.21	0.72	4109.70	17.00	0.00	262965.00	0.32	21.30	111.02	563007000.00	101.10	-0.83	0.28	30.00
3	American Samoa	ASM	Oceania	NaN	55.62	-0.05	0.87	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.20	NaN	NaN
4	Andorra	AND	Europe	711.55	77.00	-0.07	0.88	38962.61	NaN	NaN	NaN	NaN	NaN	104.89	NaN	NaN	1.42	NaN	NaN

Analyse exploratoire

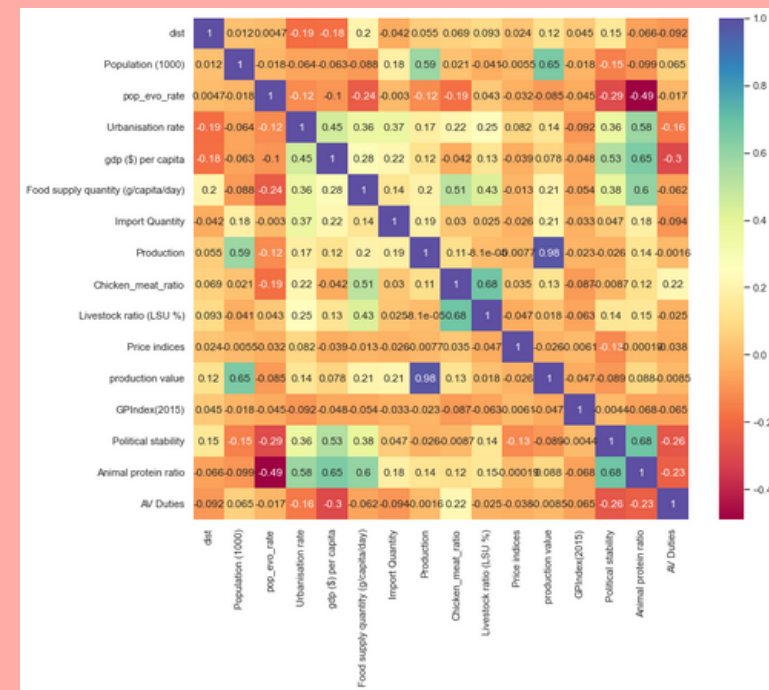
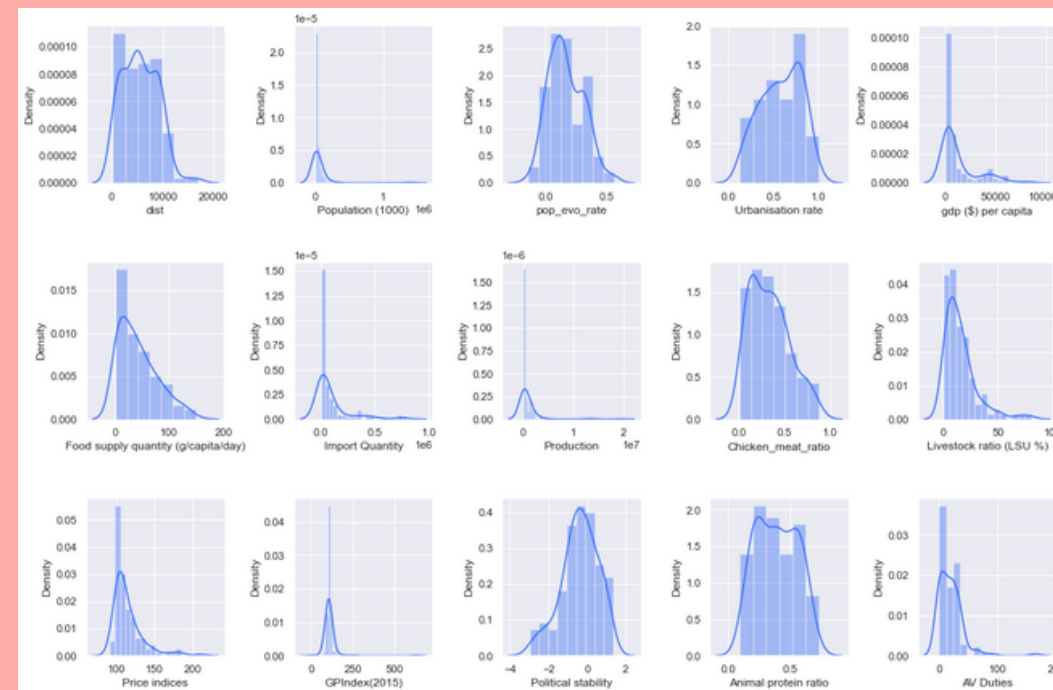
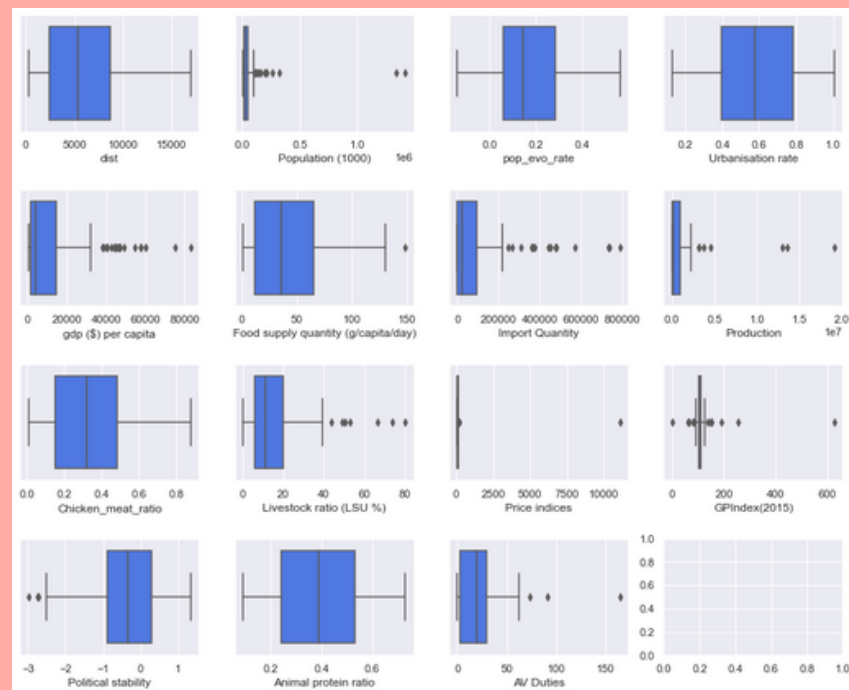
Afin de travailler sur un jeu de données cohérent, plusieurs manipulations ont été effectuées pour analyser et/ou corriger :

- les valeurs manquantes,
- les valeurs aberrantes,
- la distribution des variables,
- les relations entre nos variables

NB : La taille critique de nos marchés potentiels a été fixée à 5 millions d'habitants

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 235
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Area                                  115 non-null    object
1   alpha-3                              115 non-null    object
2   region                               115 non-null    object
3   dist                                  115 non-null    float64
4   Population (1000)                    115 non-null    float64
5   pop_evo_rate                         115 non-null    float64
6   Urbanisation rate                    115 non-null    float64
7   gdp ($) per capita                    115 non-null    float64
8   Food supply quantity (g/capita/day)  115 non-null    float64
9   Import Quantity                      115 non-null    float64
10  Production                           115 non-null    float64
11  Chicken_meat_ratio                   115 non-null    float64
12  Livestock ratio (LSU %)               115 non-null    float64
13  Price indices                        115 non-null    float64
14  GPIIndex(2015)                       115 non-null    float64
15  Political stability                   115 non-null    float64
16  Animal protein ratio                 115 non-null    float64
17  AV Duties                            115 non-null    float64
dtypes: float64(15), object(3)
memory usage: 17.1+ KB
```



4

Méthode 1 : Classification ascendante hiérarchique

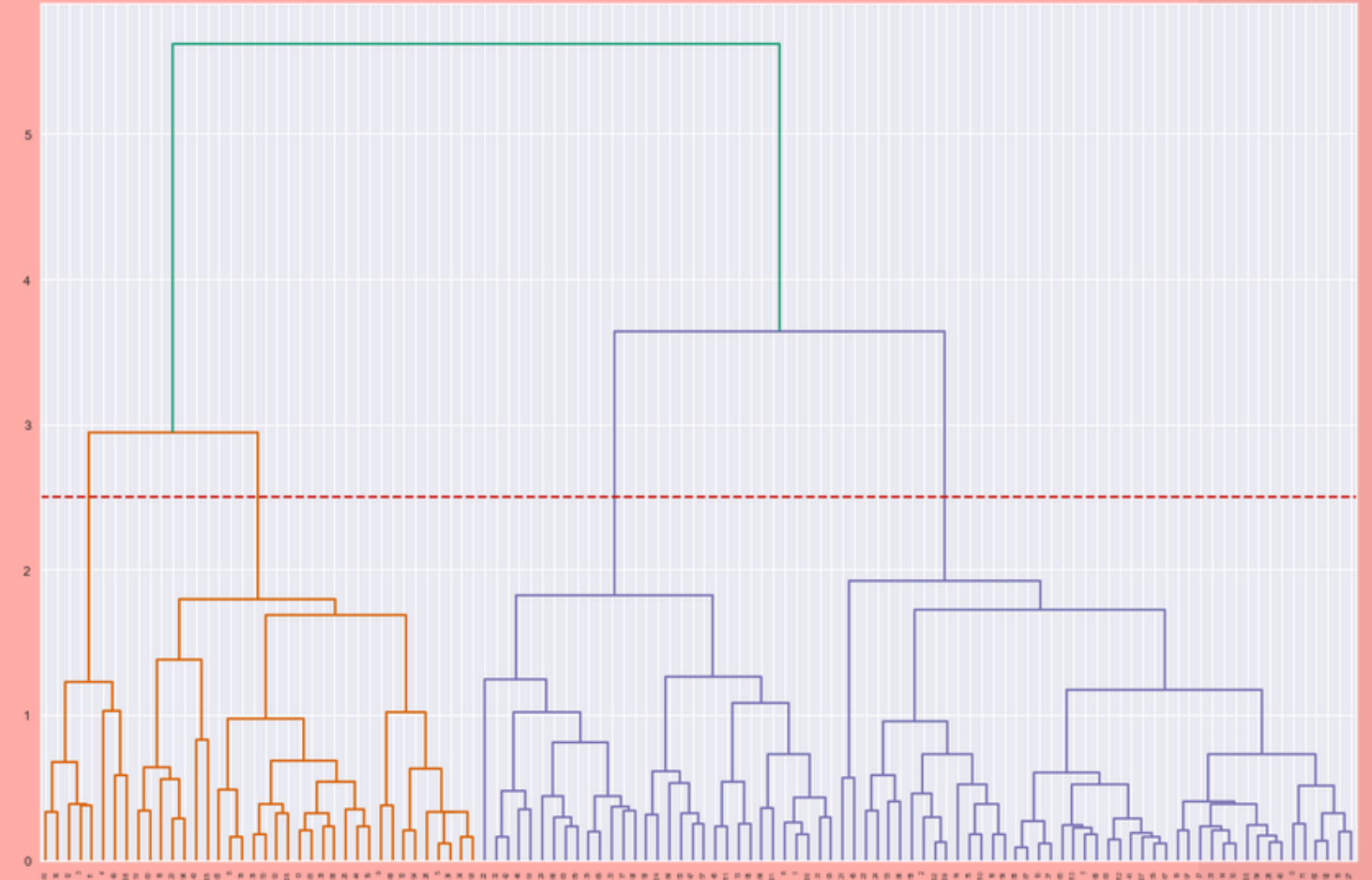
La classification ascendante hiérarchique (CAH) est une méthode de classification itérative dont le principe est de regrouper les individus (pays) les plus "proches" (par rapport à leurs variables) selon un nombre de clusters à définir.

Sa représentation graphique s'appelle un dendrogramme.

Chaque trait (bas du graphique) représente un pays.

Chaque branche correspond à différents niveaux de regroupement.

Nous allons définir 4 clusters et analyser les moyennes de nos variables retenues pour chacun de ces clusters.



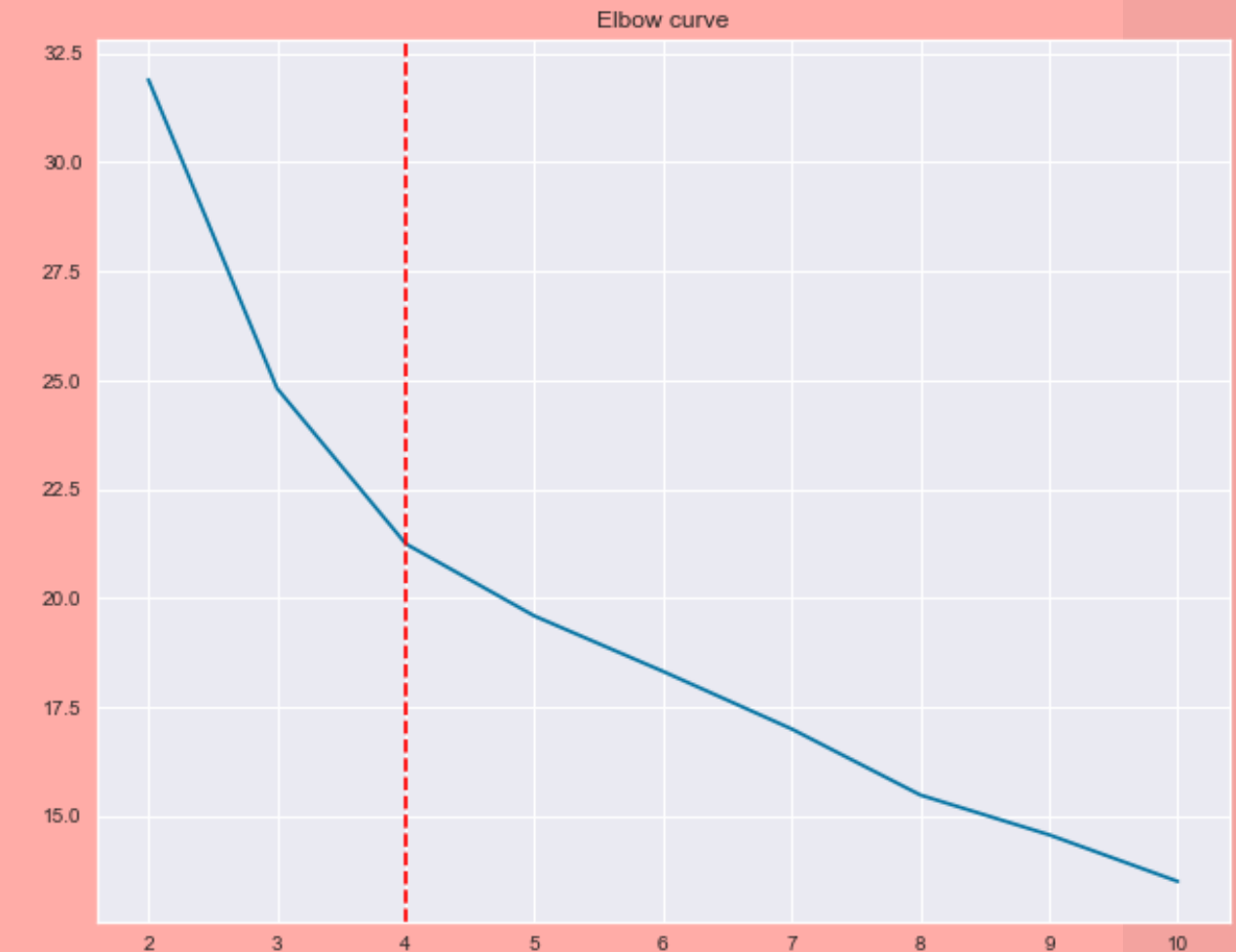
	dist	Population (1000)	gdp (\$) per capita	Food supply quantity (g/capita/day)	Chicken_meat_ratio	Political stability	Animal protein ratio	AV Duties	Production per 1000 capita	Importation per 1000 capita
Cluster_Labels										
0	6389.12	92067.75	2151.91	13.60	0.16	-0.69	0.27	24.63	2.83	2.35
1	5978.08	50516.18	4784.69	52.48	0.56	-1.05	0.35	31.34	17.62	2.92
2	9664.89	85428.25	25683.74	122.62	0.55	-0.01	0.57	7.79	53.85	0.50
3	2947.10	31144.92	34118.88	55.70	0.32	0.63	0.57	2.22	24.17	10.19

5

Méthode 2 : K-means

C'est un des algorithmes de clustering les plus répandus qui permet d'analyser un jeu de données caractérisées par un ensemble de descripteurs, afin de regrouper les données "similaires" en groupes (ou clusters).

Nous allons définir 4 clusters et analyser les moyennes de nos variables retenues pour chacun de ces clusters (par leurs centroïdes).



	dist	Population (1000)	gdp (\$) per capita	Food supply quantity (g/capita/day)	Chicken_meat_ratio	Political stability	Animal protein ratio	AV Duties	Production per 1000 capita	Importation per 1000 capita
0	6493.77	53421.61	5410.91	57.91	0.61	-0.96	0.36	34.69	19.60	3.24
1	2530.22	26895.51	35976.39	49.56	0.29	0.64	0.58	2.69	20.42	11.03
2	6126.98	88786.84	1937.77	13.96	0.19	-0.84	0.26	24.86	3.24	2.09
3	8069.00	65735.15	19127.07	105.40	0.51	0.09	0.54	8.67	45.66	2.16

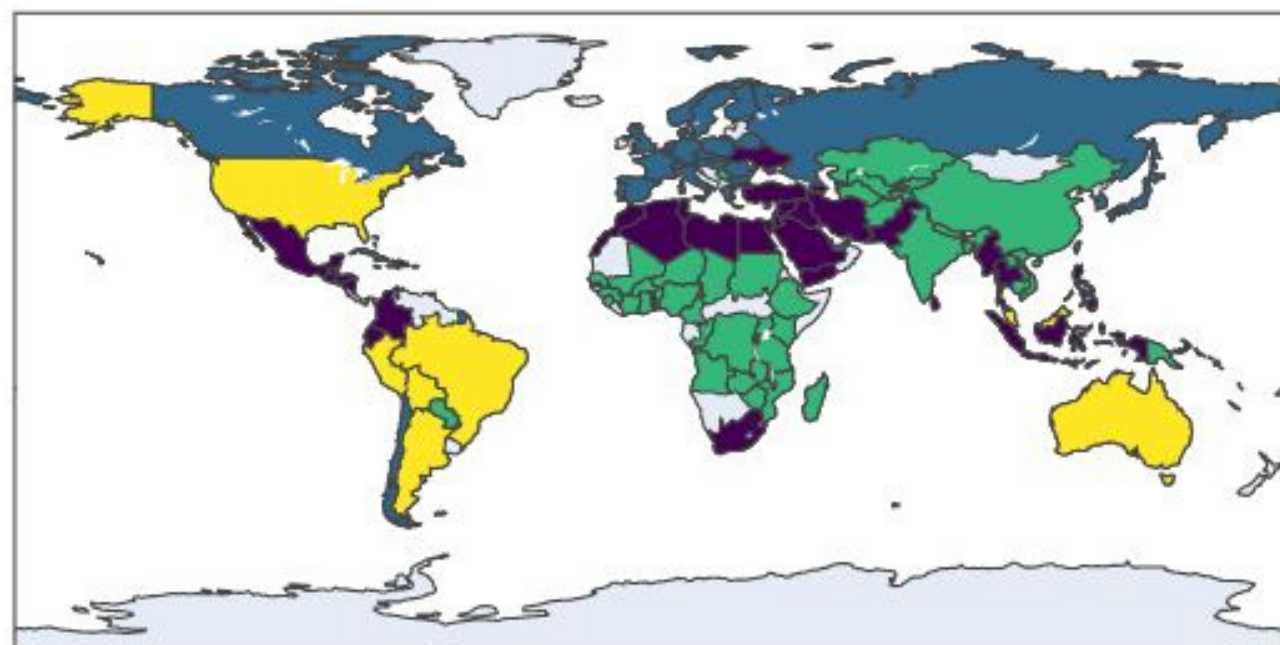
6

Comparaison des 2 méthodes

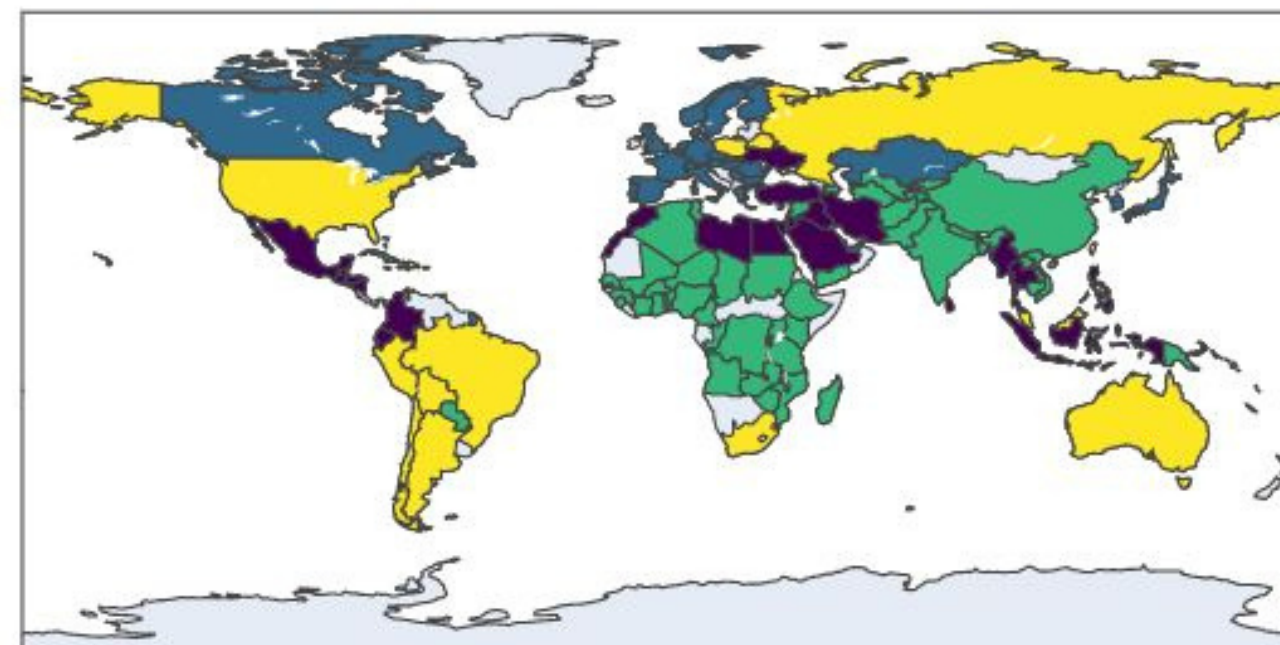
L'utilisation de 2 méthodes différentes nous permet de vérifier la cohérence des clusters obtenus.

Visualisation de nos pays cibles selon leurs clusters pour chacune de nos méthodes :

ACH clusters



K_means clusters



Hormis quelques exceptions (Russie, Algérie notamment), les 2 modèles de clustering donnent des résultats similaires. Les résultats de la méthode K-means seront retenus pour la suite de notre analyse.

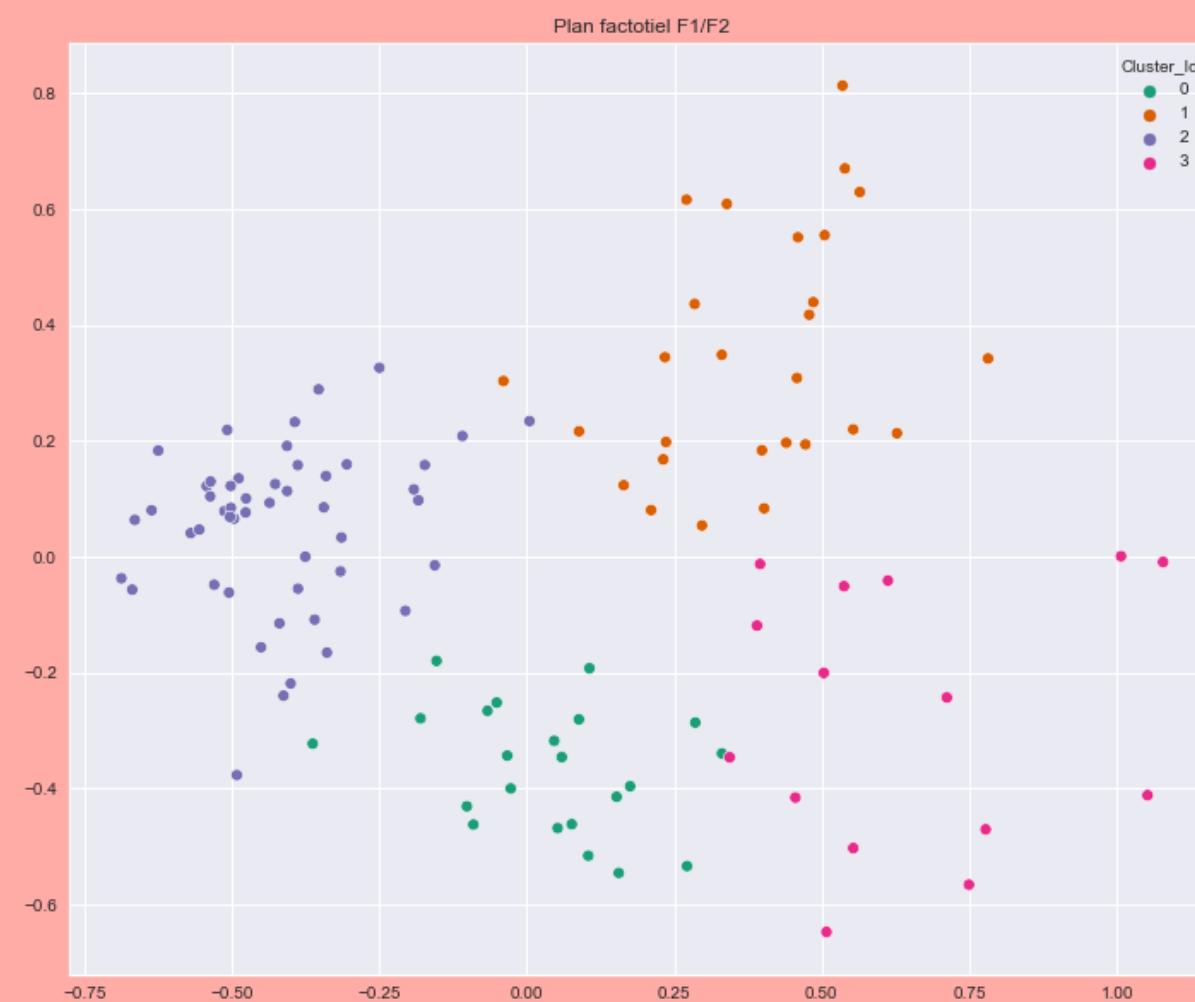
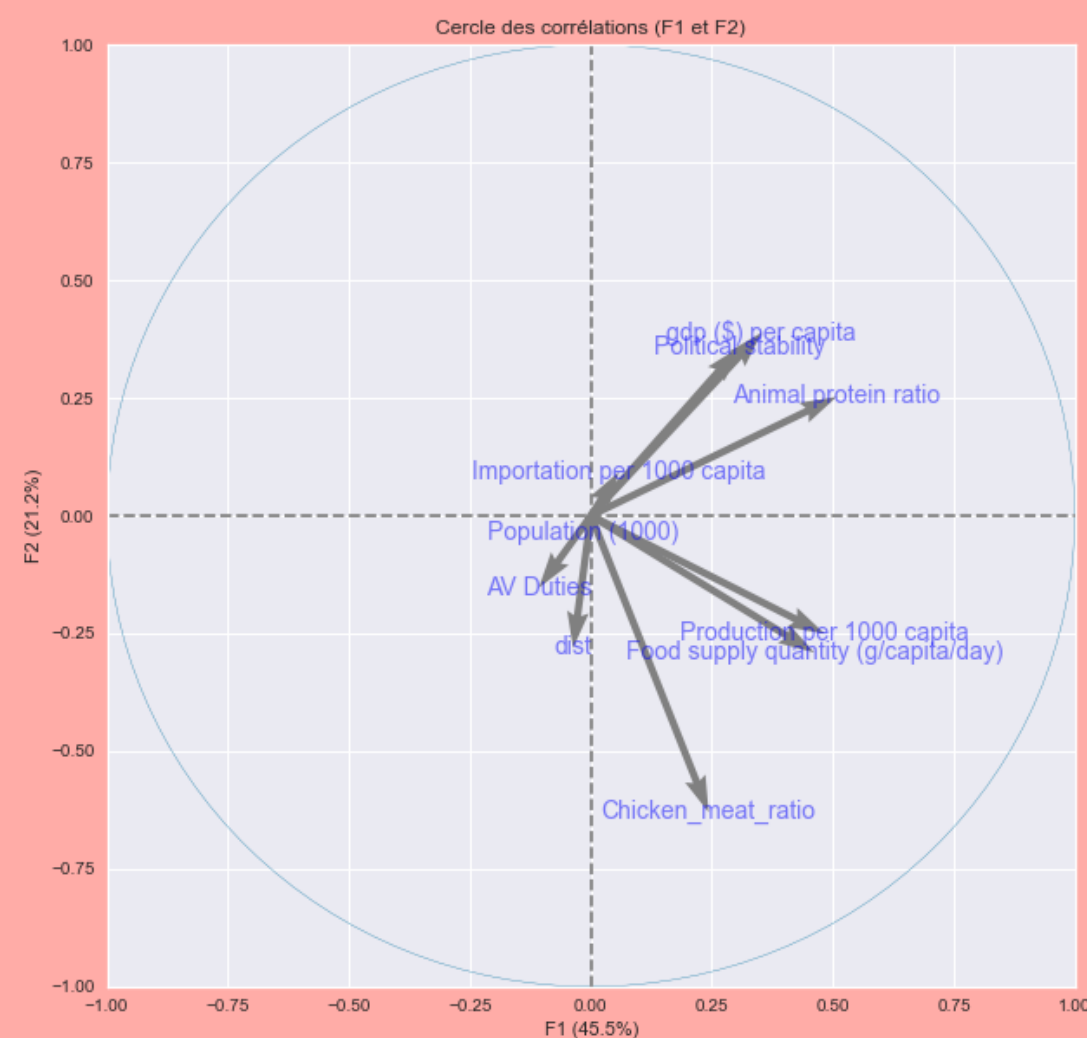
7

Analyse en composantes principales

Technique statistique permettant la réduction de dimension (le nombre de variables) afin d'analyser :

- la variabilité de nos individus (pays)
- les relations entre nos variables.

Cela facilite également la représentation graphique de nos données.



Cercle de corrélation (liaison de nos variables)

La représentativité des variables sur F1/F2 est plutôt moyenne (loin du cercle de corrélation). 67% de la variance est expliquée par les 2 1ères composantes principales.

- Les variables Animal protein ratio, Production per 1000 capita et Food supply quantity sont moyennement corrélées à F1 (environ 0,50).
- Chicken meat ratio est plus fortement corrélée à F2 (0,6).

Plan factoriel (Variabilité de nos pays)

Le clustering semble être efficace (distinction nette des groupes de pays)

Recommandations

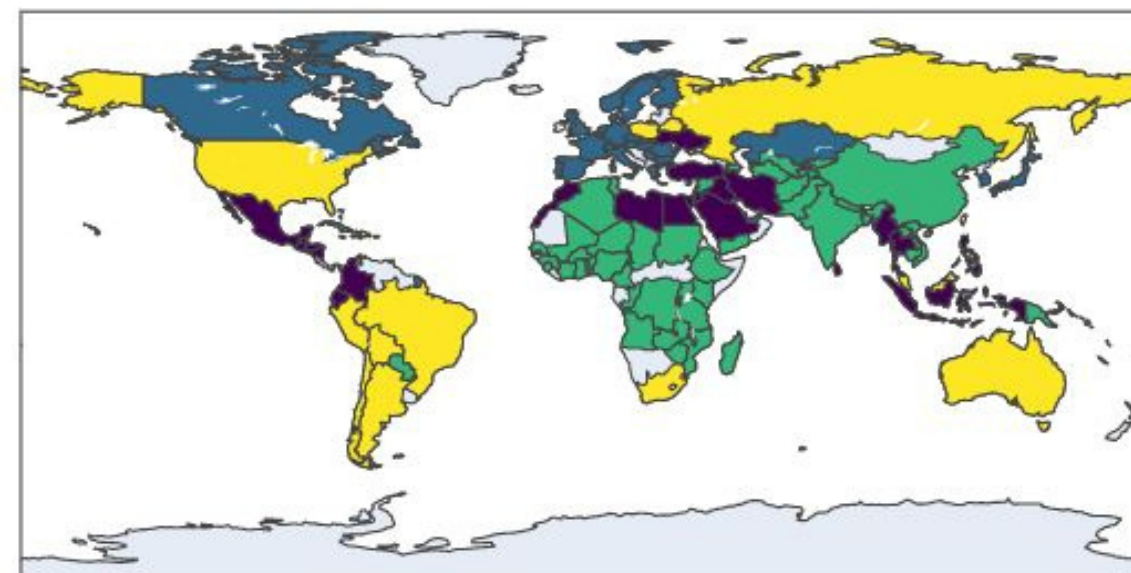
	dist	Population (1000)	gdp (\$) per capita	Food supply quantity (g/capita/day)	Chicken_meat_ratio	Political stability	Animal protein ratio	AV Duties	Production per 1000 capita	Importation per 1000 capita
0	6493.77	53421.61	5410.91	57.91	0.61	-0.96	0.36	34.69	19.60	3.24
1	2530.22	26895.51	35976.39	49.56	0.29	0.64	0.58	2.69	20.42	11.03
2	6126.98	88786.84	1937.77	13.96	0.19	-0.84	0.26	24.86	3.24	2.09
3	8069.00	65735.15	19127.07	105.40	0.51	0.09	0.54	8.67	45.66	2.16

Le **cluster 1** (couleur bleue) semble plutôt intéressant compte tenu des caractéristiques suivantes :

- Cluster de pays les moins éloignés
- Taille du marché raisonnable
- PIB par habitant le plus élevé
- Disponibilité alimentaire en viande de poulets moyenne
- 1/3 de la viande consommée est du poulet
- Forte stabilité politique
- Plus de la moitié des protéines animales consommées sont d'origine animale
- Très faibles taxes d'importation
- Gros importateur de viande de poulets - les importations correspondent à la moitié de la production nationale

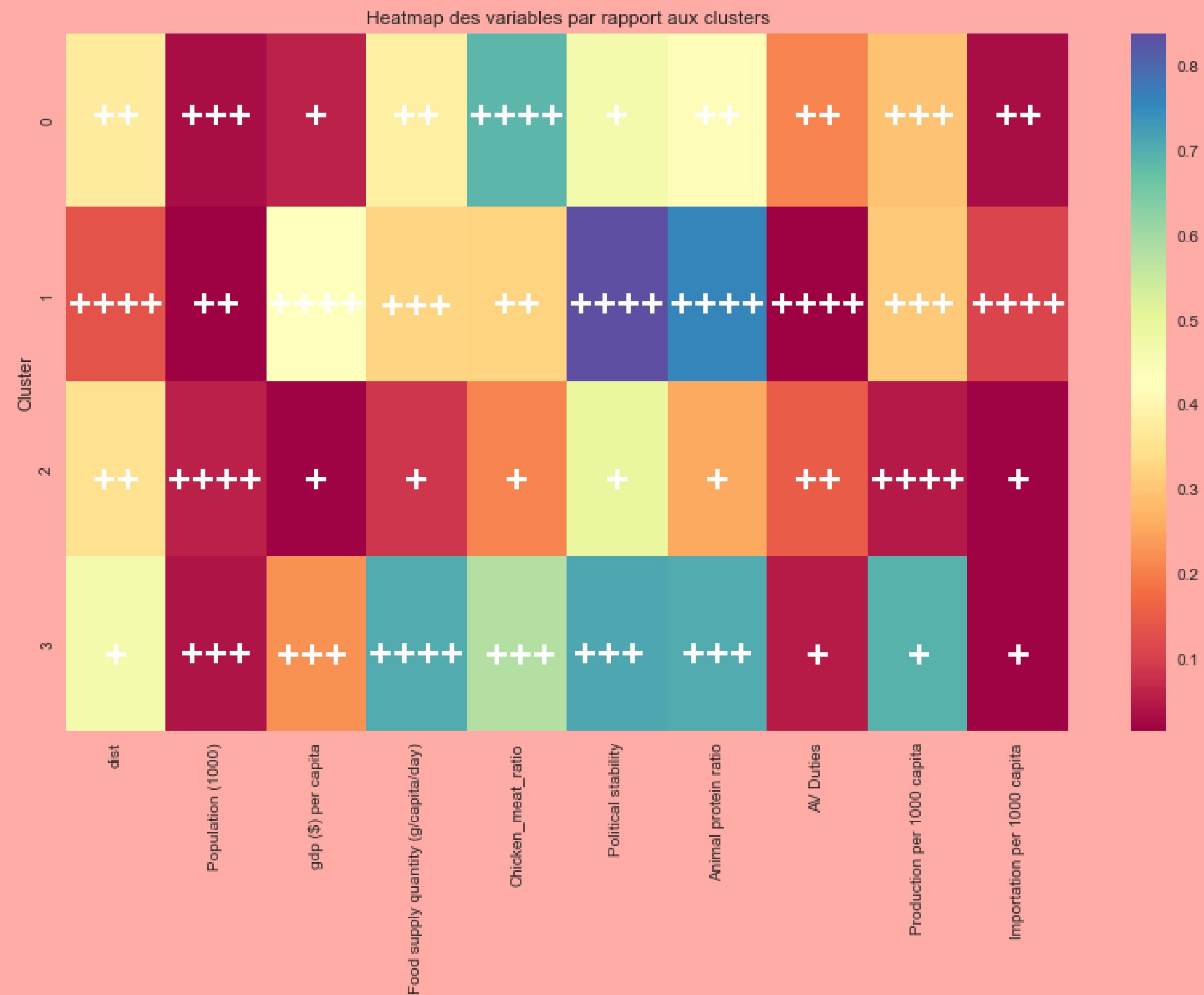
Les informations sont faites sur la base des moyennes observées.

K_means clusters



Pays à fort potentiel :

Autriche, Belgique, Bulgarie, République Tchèque, Danemark, Finlande, Allemagne, Grèce, Hong Kong, Hongrie, Italie, Japon, Kazakhstan, Pays Bas, Norvège, Portugal, Corée du Sud, Roumanie, Serbie, Slovaquie, Espagne, Suède, Suisse, Emirats Arabes Unis, UK



Ce schéma permet d'analyser la qualité de chacun de nos clusters au regard de chacune des variables retenues.

Celui-ci confirme le **fort potentiel du cluster 1** pour le développement à l'international de notre activité.

NB : liste des pays par cluster à la page suivante.

NB2 : ++++ indicateur très favorable

+ indicateur limitant

Cluster 0 - 22 countries

'Colombia', 'Ecuador', 'Egypt', 'El Salvador', 'Guatemala', 'Honduras', 'Indonesia', 'Iran (Islamic Republic of)', 'Iraq', 'Jordan', 'Lebanon', 'Libya', 'Mexico', 'Morocco', 'Myanmar', 'Nicaragua', 'Philippines', 'Saudi Arabia', 'Sri Lanka', 'Thailand', 'Turkey', 'Ukraine'

Cluster 1 - 27 countries

'Austria', 'Belgium', 'Bulgaria', 'Canada', 'Czechia', 'Denmark', 'Finland', 'France', 'Germany', 'Greece', 'Hong Kong', 'Hungary', 'Italy', 'Japan', 'Kazakhstan', 'Netherlands', 'Norway', 'Portugal', 'Republic of Korea', 'Romania', 'Serbia', 'Slovakia', 'Spain', 'Sweden', 'Switzerland', 'United Arab Emirates', 'United Kingdom of Great Britain and Northern Ireland'

Cluster 2 - 50 countries

'Afghanistan', 'Algeria', 'Angola', 'Azerbaijan', 'Bangladesh', 'Benin', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon', 'Chad', 'China', 'Congo', 'Cuba', 'Côte d'Ivoire', 'Democratic Republic of the Congo', 'Ethiopia', 'Ghana', 'Guinea', 'Haiti', 'India', 'Kenya', 'Kyrgyzstan', 'Lao People's Democratic Republic', 'Madagascar', 'Malawi', 'Mali', 'Mozambique', 'Nepal', 'Niger', 'Nigeria', 'Pakistan', 'Papua New Guinea', 'Paraguay', 'Rwanda', 'Senegal', 'Sierra Leone', 'Sudan', 'Syrian Arab Republic', 'Tajikistan', 'Togo', 'Tunisia', 'Turkmenistan', 'Uganda', 'United Republic of Tanzania', 'Uzbekistan', 'Viet Nam', 'Yemen', 'Zambia', 'Zimbabwe'

Cluster 3 - 15 countries

'Argentina', 'Australia', 'Belarus', 'Bolivia (Plurinational State of)', 'Brazil', 'Chile', 'Dominican Republic', 'Israel', 'Malaysia', 'Peru', 'Poland', 'Russian Federation', 'South Africa', 'Taiwan', 'United States of America'