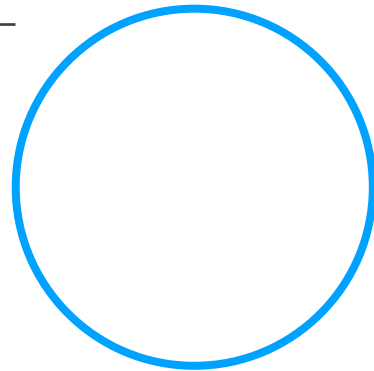




Étude de marché internationale

MAXIMEBCH – DATA ANALYST



INTRODUCTION

- Entreprise française de volaille souhaitant exporter ses produits à l'international
- Objectif : Réalisation d'une étude de marché pour identifier les pays à cibler





Sommaire

- 1 – NETTOYAGE DES DONNÉES
- 2 – ANALYSE DES DONNÉES
- 4 – CONCLUSIONS ET RECOMMANDATIONS

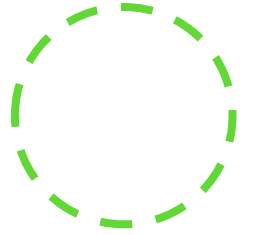


1 - NETTOYAGE DES DONNÉES



1 – NETTOYAGE DES DONNÉES

FICHIERS SOURCES

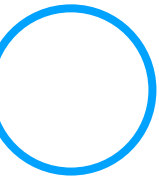


4 fichiers .csv extraits des données de la FAO :

- Bilan alimentaire : disponibilités alimentaires en calories et en protéines, d'origine animale ou non
- Marché du poulet : production, importation et exportation de poulet
- Population : chiffres de 2008 et 2018
- Sécurité et stabilité : PIB par habitant, indice de stabilité politique et d'absence de violence/terrorisme

Les fichiers ne présentent pas de doublons ou de valeurs manquantes.

On supprime la zone « Chine » qui est un agrégat de la Chine continentale, Taïwan, Hong Kong et Macao.



1 – NETTOYAGE DES DONNÉES

CRÉATION D'UN SEUL DATAFRAME

Après concaténation et pivot, les données sont rassemblées par pays dans un seul dataframe.

De nouvelles variables ont été créées :

- Ratio_protéines_animales = protéines animales / total protéines
- Population_croissance = population 2018 / population 2008
- Ppa (PIB par habitant) = PIB / population
- PIB_croissance = PIB 2018 / PIB 2008
- Poulet_import-export = importations 2018 / exportations 2018

pays	dispo_calories	dispo_proteines	ratio_proteines_animales	population	population_croissance	pib	ppa	pib_croissance	poulet_import-export
Afghanistan	2040.0	55.52	0.194344	37.171921	1.340868	2190.2	0.000059	1.379393	155.279221
Albania	3360.0	115.74	0.533523	2.882740	0.960056	13601.3	0.004718	1.344122	NaN
Algeria	3322.0	91.83	0.269302	42.228408	1.215885	11479.5	0.000272	1.080424	NaN
American Samoa	NaN	NaN	NaN	0.055465	0.964743	NaN	NaN	NaN	NaN
Andorra	NaN	NaN	NaN	0.077006	0.918247	NaN	NaN	NaN	NaN

1 – NETTOYAGE DES DONNÉES

CRÉATION DU DATAFRAME

Après suppression des valeurs manquantes et nulles :

- 106 pays
- 83% de la population mondiale


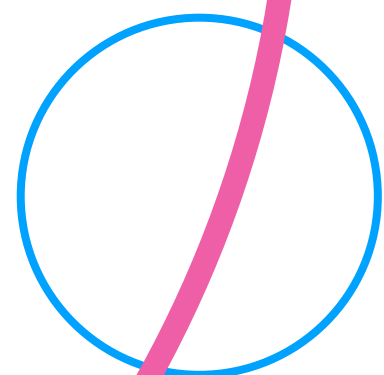

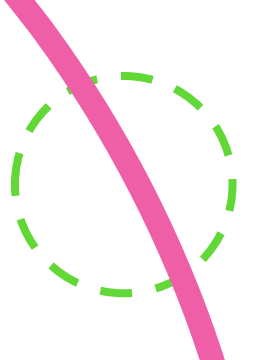
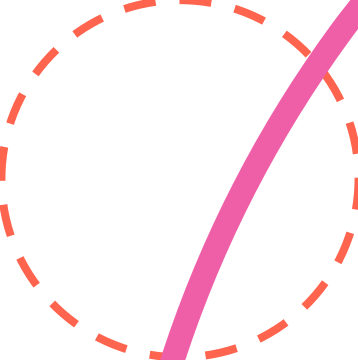

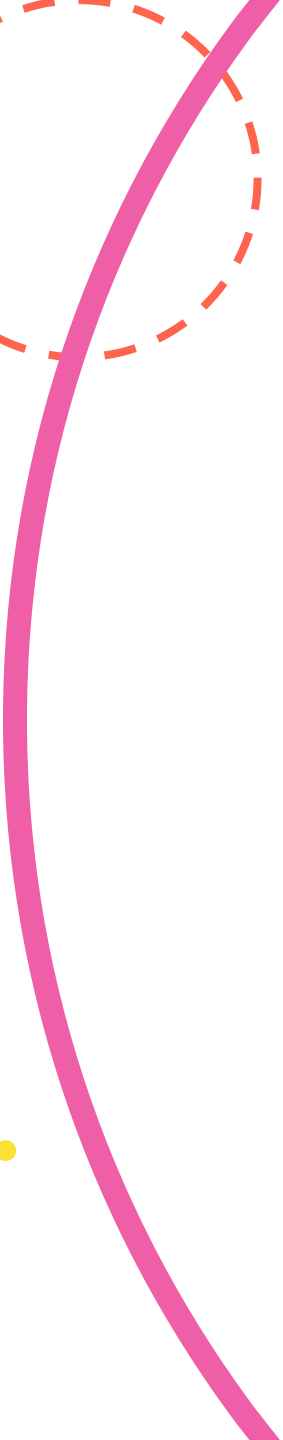
	pays	dispo_calories	dispo_proteines	ratio_proteines_animales	population	population_croissance	plb	plb_h	plb_croissance	poulet_Import-export
0	Afrique du Sud	2899.000000	84.370000	0.427048	57.792518	1.160971	12627.900000	0.000219	0.999978	10.729167
1	Allemagne	3554.000000	105.400000	0.607495	83.124418	1.025395	53483.400000	0.000643	1.123618	1.360937
2	Arabie saoudite	3307.000000	86.940000	0.376582	33.702758	1.301841	47592.800000	0.001412	1.049684	18.285714
3	Argentine	3307.000000	106.770000	0.646717	44.381150	1.108811	22759.400000	0.000513	0.985106	0.078788
4	Australie	3391.000000	105.940000	0.667359	24.898152	1.167158	49151.700000	0.001974	1.096270	0.410256
5	Autriche	3695.000000	109.120000	0.603281	8.891388	1.065918	55297.600000	0.006219	1.041584	1.373494
6	Azerbaïdjan	3149.000000	94.420000	0.344524	9.949537	1.127826	14209.600000	0.001428	1.101246	35.000000
7	Belgique	3769.000000	99.870000	0.586563	11.482178	1.065260	51111.000000	0.004451	1.055520	0.474178
	Bolivie (État									

1 – NETTOYAGE DES DONNÉES

GEOPLOT

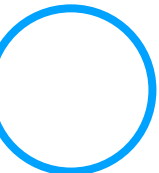
- GEOPLOT est une librairie Python permettant le traçage géospatial
- Jointure avec un DF contenant le continent et la géolocalisation de chaque pays

	continent	pays	iso_a3	geometry	dispo_calories	dispo_proteines	ratio_proteines_animaux	population	population_croissance	pib
0	Oceania	Fiji	FJI	MULTIPOLYGON (((180.00000 -16.06713, 180.00000...	2781.0	71.14	0.399635	0.883483	1.045096	13808.1
1	Africa	United Republic of Tanzania	TZA	POLYGON ((33.90371 -0.95000, 34.07262 -1.05982...	2373.0	58.93	0.206007	56.313438	1.345475	2590.0
2	North America	Canada	CAN	MULTIPOLYGON (((122.84000 49.00000, -122.9742...	3566.0	104.12	0.545044	37.074562	1.112093	48924.4



2 – CLASSIFICATION HIÉRARCHIQUE

- Entre la disponibilité en protéines et en calories
- Entre le PIB (niveau de richesse) et disponibilité alimentaire et ratio de protéines animales





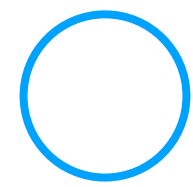
1 – CLASSIFICATION HIÉRARCHIQUE

NORMALISATION

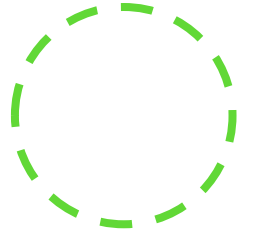


Normalisation de variables (entre 0 et 1)

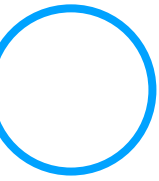
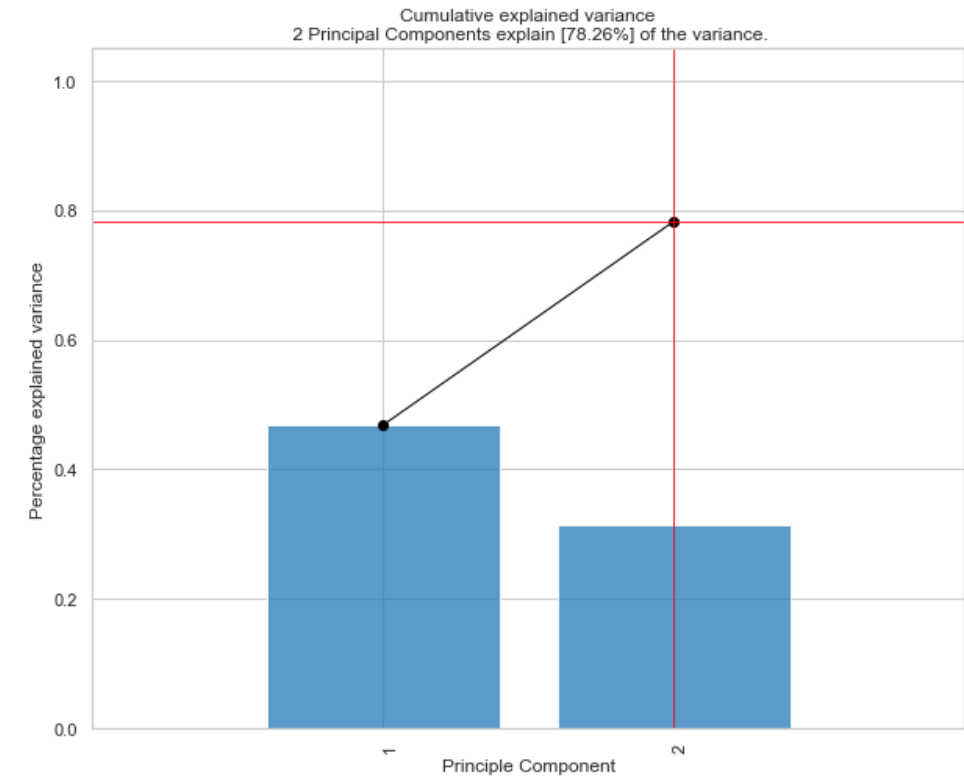
	ratio_proteines_animaux	population_croissance	pib_croissance	dispo_calories	dispo_proteines	population	pib	ppa	poulet_impo exp
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	1.000000e+02	100.0000
mean	0.454784	1.132678	1.197033	0.227394	0.006225	0.005891	0.941274	1.907658e-07	0.0309
std	0.143988	0.152900	0.218673	0.192240	0.005115	0.021517	0.105064	3.599276e-07	0.1157
min	0.141283	0.871890	0.691271	0.030351	0.000952	0.000005	0.421799	6.761622e-10	0.0000
25%	0.347529	1.029482	1.051039	0.082764	0.002498	0.000214	0.953808	1.930648e-08	0.0000
50%	0.487820	1.112964	1.185242	0.158154	0.004274	0.000761	0.987013	7.515395e-08	0.0001
75%	0.568652	1.198523	1.302479	0.280879	0.007577	0.003388	0.996512	1.837584e-07	0.0026
max	0.680987	1.755557	2.056280	0.905994	0.029119	0.190422	0.999539	2.266971e-06	0.7789



1 – CLASSIFICATION HIÉRARCHIQUE ANALYSE EN COMPOSANTE PRINCIPALE

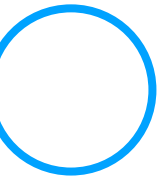
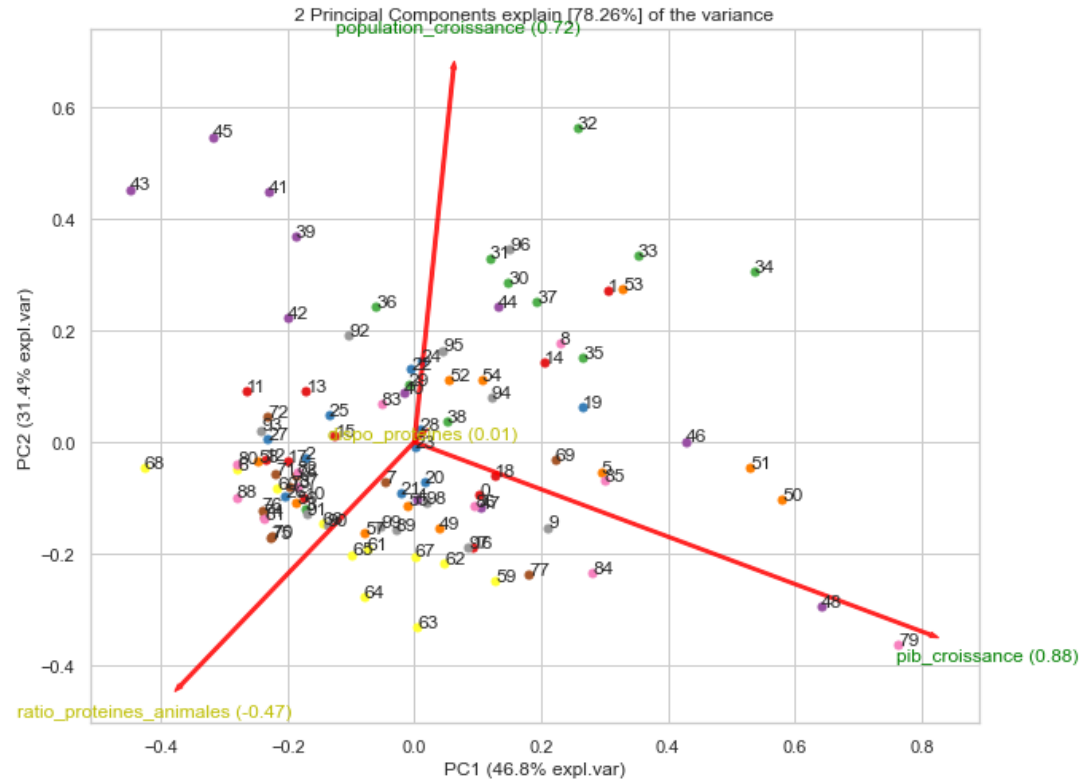
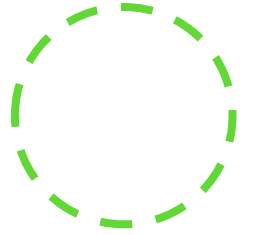


- Les variables catégorielles et la disponibilité en calories sont écartées
- Règle du coude : 2 composantes expliquent 78% de la variance

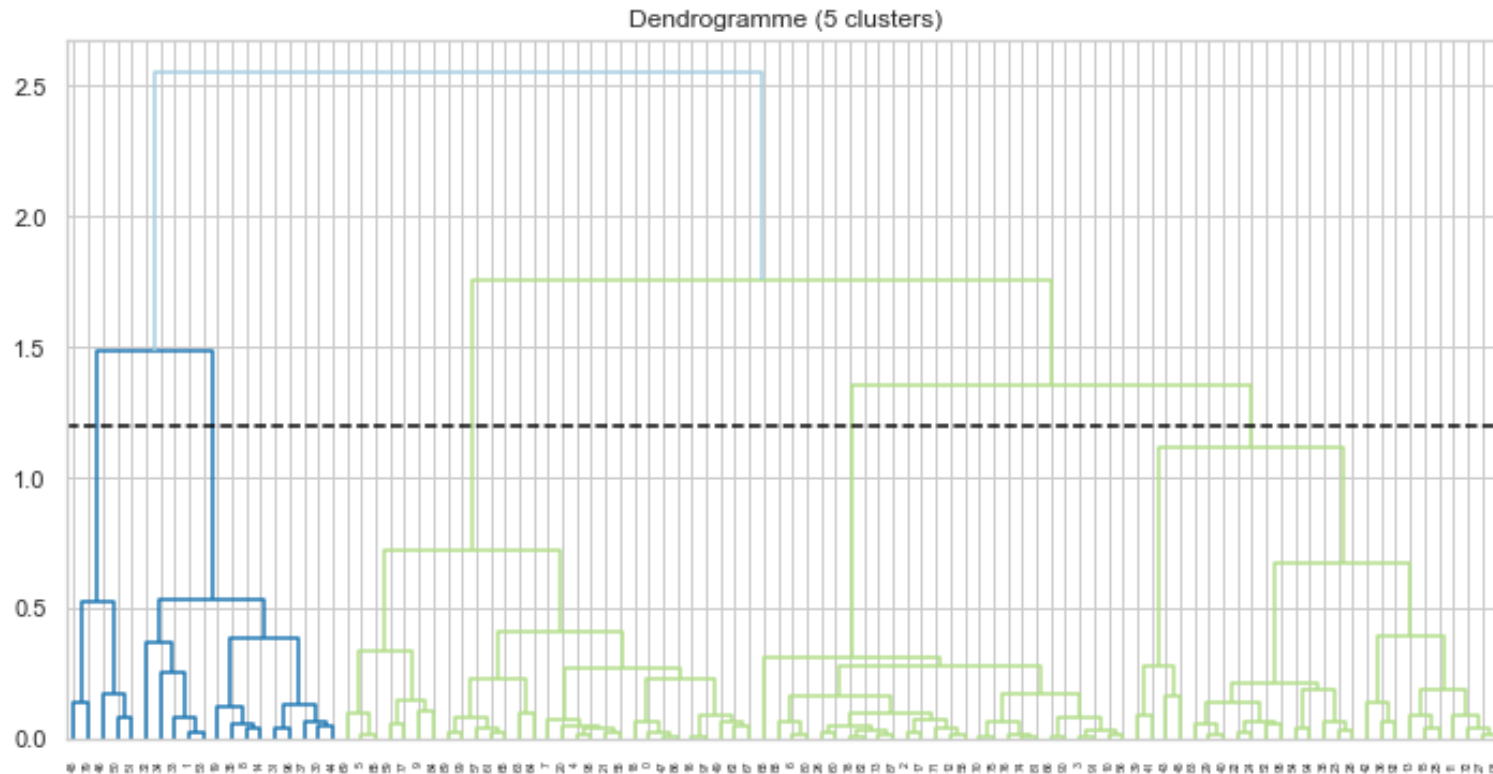


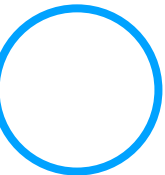


1 – CLASSIFICATION HIÉRARCHIQUE REPRÉSENTATION EN 2 DIMENSIONS



1 – CLASSIFICATION HIÉRARCHIQUE CLUSTERS – DENDROGRAMME SUR LES 2 COMPOSANTS







1 – CLASSIFICATION HIÉRARCHIQUE CLUSTERS



```
# Pays dans chaque cluster
for i in df.sort_values('cluster')['cluster'].unique():
    print('Cluster', i, ': ', len(df[df['cluster']==i]), 'pays\n' +
          ', '.join(df[df['cluster']==i]['pays'].values), '\n' + '-' * 3)
```

Cluster 0 : 26 pays

Bahamas, South Africa, Mexico, Honduras, El Salvador, Guatemala, Suriname, Jamaica, Botswana, Namibia, Congo, Eswatini, Lebanon, Tunisia, Jordan, United Arab Emirates, Kuwait, Oman, Pakistan, Kyrgyzstan, Luxembourg, Azerbaijan, Saudi Arabia, Cyprus, Morocco, Egypt

- - -

Cluster 1 : 14 pays

United Republic of Tanzania, Kenya, Lesotho, Peru, Senegal, Benin, Niger, Togo, Ghana, Côte d'Ivoire, Zambia, Iraq, Afghanistan, Uganda

- - -

Cluster 2 : 29 pays

Fiji, Kazakhstan, Indonesia, Chile, Dominican Republic, Uruguay, Bolivia (Plurinational State of), Colombia, Costa Rica, Thailand, Republic of Korea, Armenia, Belarus, Poland, Hungary, Romania, Lithuania, Latvia, Estonia, Bulgaria, Turkey, Ireland, Georgia, Philippines, Malaysia, Slovakia, Bosnia and Herzegovina, North Macedonia, Serbia

- - -

Cluster 3 : 5 pays

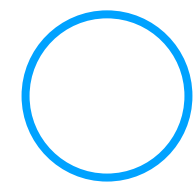
Cambodia, Myanmar, India, Bangladesh, China, mainland

- - -

Cluster 4 : 26 pays

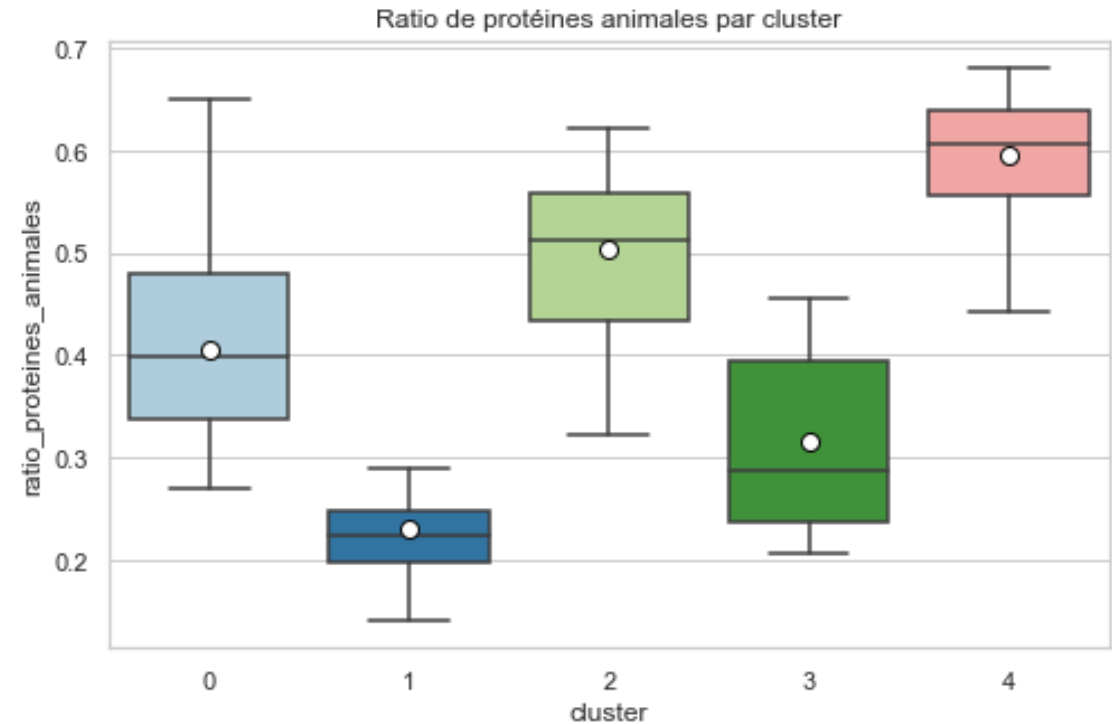
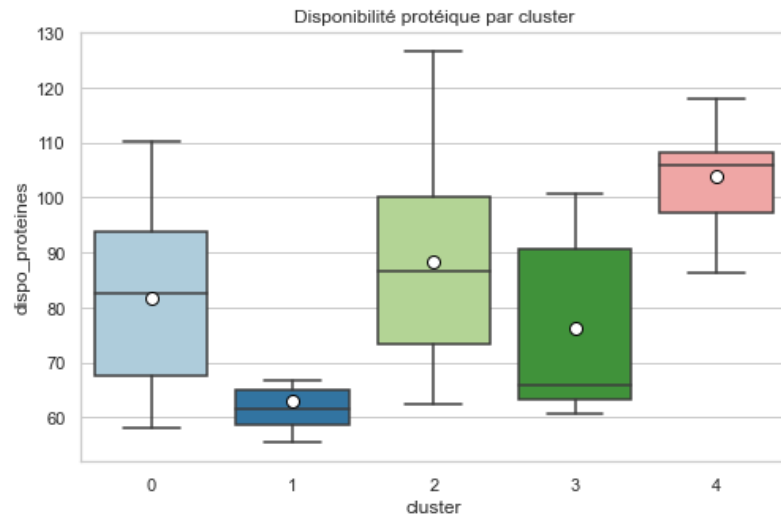
Canada, United States of America, Argentina, Russian Federation, Norway, Brazil, France, Sweden, Ukraine, Austria, Germany, Greece, Croatia, Switzerland, Belgium, Netherlands, Portugal, Spain, Australia, Italy, Denmark, United Kingdom of Great Britain and Northern Ireland, Slovenia, Finland, Czechia, Japan

- - -



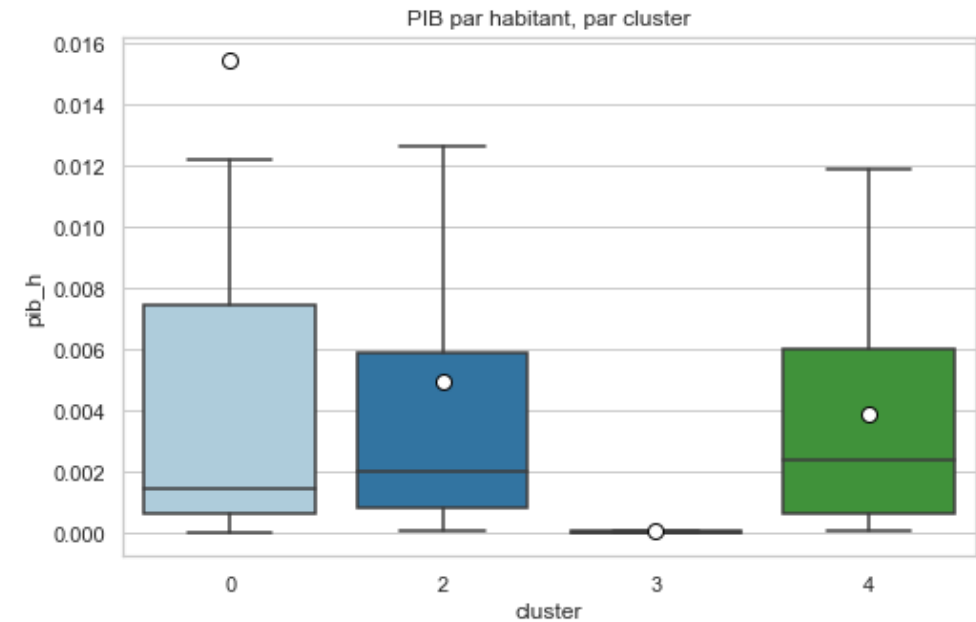
1 – CLASSIFICATION HIÉRARCHIQUE

- Les pays du cluster 2 consomment moins de protéines et de protéines animales, il faut l'écarter.

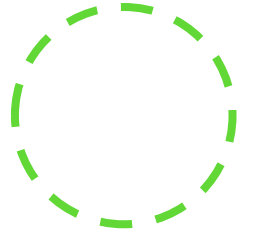


1 – CLASSIFICATION HIÉRARCHIQUE

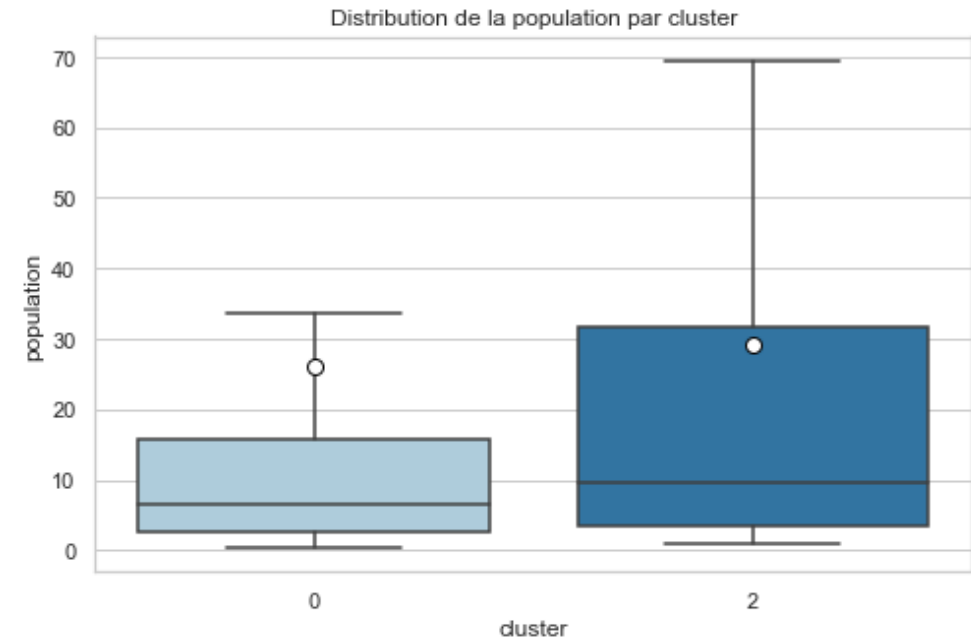
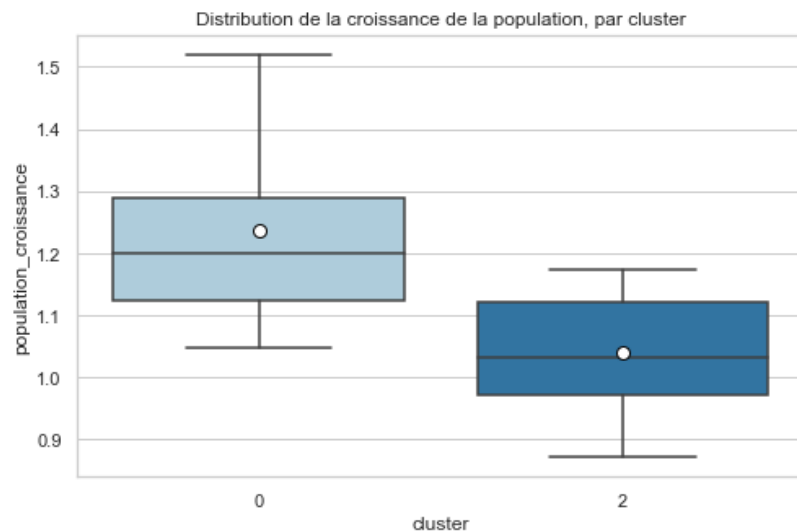
- Le cluster 4 a une forte croissance de PIB mais le PIB et le PIB par habitant restent faibles. Il convient donc de l'écarter.



1 – CLASSIFICATION HIÉRARCHIQUE



- Le cluster 1 a une plus faible population mais sa croissance est forte. Il convient de le sélectionner car son importation de poulet est beaucoup plus forte que celle du 3.



1 – CLASSIFICATION HIÉRARCHIQUE

Cluster 1 par ratio d'import de |



Classement du cluster 1 par ratio de protéines animales

