



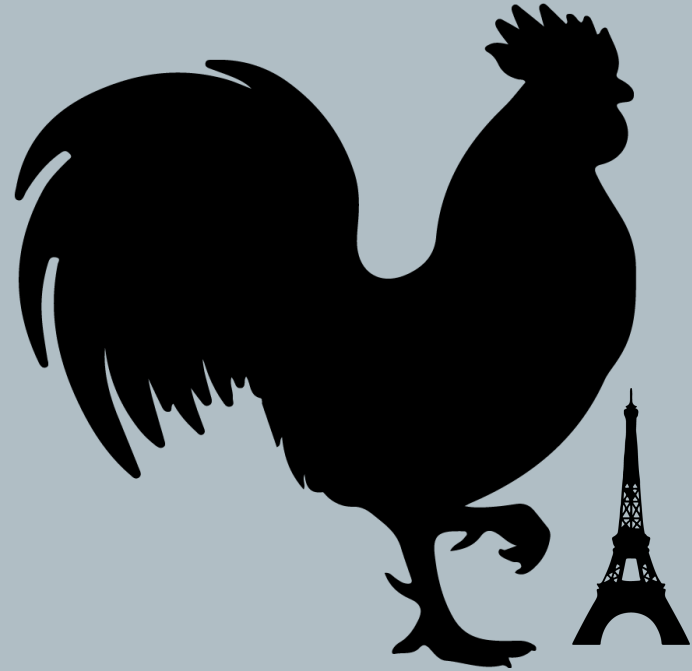
Parcours Data Analyst : projet 5

Étude de marché

Sélection de pays pour l'exportation de poulet

- Entreprise française
- Spécialisée dans le poulet
- Envie de développement à l'international
- Exportation uniquement

Cibler les pays propices à une insertion
sur le marché du poulet



1. Préparation des données

- Données
- Concaténation et pivot
- Variables
- Valeurs manquantes
- Geoplot
- 100 pays

2. Classification hiérarchique

- Corrélations
- Centrage et normalisation
- Analyse en composantes principales
- Représentation en 2 dimensions
- Dendrogramme
- Visualisation des clusters
- Tests statistiques

3. Analyse des groupes

- Carte mondiale des clusters
- Indicateurs protéiques
- Indicateurs de PIB
- Import / export de poulet
- Indicateurs démographiques
- Résumé des groupes
- Groupe choisi
- Analyse du groupe
- Sous-groupe final

Préparation des données

1

4 datasets :

Source : **FAO**

1. **Bilans alimentaires**
Disponibilités en calories et protéines
Distinction par produits d'origine animale
2. **Chiffres du poulet**
Production, importation, exportation
3. **Population**
2008 et 2018
4. **Indicateurs de PIB**
PIB, PIB par habitant

	Domain	Area	Element	Item	Year	Unit	Value	Note
	Domain	Area	Element	Item	Year	Unit	Value	Note
0	Suite of Food Security Indicators	Afghanistan	Value	Gross domestic product per capita, PPP, dissem...	2008	IS	1587.80	NaN
1	Suite of Food Security Indicators	Afghanistan	Value	Gross domestic product per capita, PPP, dissem...	2018	IS	2190.20	NaN
2	Suite of Food Security Indicators	Afghanistan	Value	Political stability and absence of violence/te...	2008	index	-2.69	NaN
3	Suite of Food Security Indicators	Afghanistan	Value	Political stability and absence of violence/te...	2018	index	-2.75	NaN
4	Suite of Food Security Indicators	Albania	Value	Gross domestic product per capita, PPP, dissem...	2008	IS	10119.10	NaN
...
759	Suite of Food Security Indicators	Zambia	Value	Political stability and absence of violence/te...	2018	index	0.14	NaN
760	Suite of Food Security Indicators	Zimbabwe	Value	Gross domestic product per capita, PPP, dissem...	2008	IS	1739.20	NaN
761	Suite of Food Security Indicators	Zimbabwe	Value	Gross domestic product per capita, PPP, dissem...	2018	IS	3130.00	NaN
762	Suite of Food Security Indicators	Zimbabwe	Value	Political stability and absence of violence/te...	2008	index	-1.21	NaN
763	Suite of Food Security Indicators	Zimbabwe	Value	Political stability and absence of violence/te...	2018	index	-0.71	NaN
764 rows x 8 columns								
468 rows x 8 columns								
549 rows x 7 columns								

- Aucune valeur manquante
- Aucun doublon

Concaténation

2473 lignes au total

	Domain	Area Code	Area	Element	Item	Year	Unit	Value	Note
0	New Food Balances	2	Afghanistan	Food supply (kcal/capita/day)	Grand Total	2018	kcal/capita/day	2040.00	NaN
1	New Food Balances	2	Afghanistan	Protein supply quantity (g/capita/day)	Grand Total	2018	g/capita/day	55.52	NaN
2	New Food Balances	2	Afghanistan	Food supply (kcal/capita/day)	Animal Products	2018	kcal/capita/day	191.00	NaN
3	New Food Balances	2	Afghanistan	Protein supply quantity (g/capita/day)	Animal Products	2018	g/capita/day	10.79	NaN
4	New Food Balances	3	Albania	Food supply (kcal/capita/day)	Grand Total	2018	kcal/capita/day	3360.00	NaN
...
759	Suite of Food Security Indicators	251	Zambia	Value	Political stability and absence of violence/te...	2018	index	0.14	NaN
760	Suite of Food Security Indicators	181	Zimbabwe	Value	Gross domestic product per capita, PPP, dissem...	2008	IS	1739.20	NaN
761	Suite of Food Security Indicators	181	Zimbabwe	Value	Gross domestic product per capita, PPP, dissem...	2018	IS	3130.00	NaN
762	Suite of Food Security Indicators	181	Zimbabwe	Value	Political stability and absence of violence/te...	2008	index	-1.21	NaN
763	Suite of Food Security Indicators	181	Zimbabwe	Value	Political stability and absence of violence/te...	2018	index	-0.71	NaN

2473 rows x 9 columns

Table pivot

- 238 pays
- Clé primaire : 'area'

Element	Area	Export Quantity	Food supply (kg/capita/day)			Import Quantity	Production	Protein supply quantity (g/capita/day)					Total Population - Both sexes		Gross domestic product per capita, PPP, dissemination (constant 2011 international \$)	Political stability and absence of violence/terrorism (index)	Value
Item		Meat, chicken	Animal Products	Grand Total	Meat, chicken	Meat, chicken	Animal Products	Grand Total	Meat, chicken	Population - Est. & Proj.							
Year		2018	2018	2018	2018	2018	2018	2018	2018	2008	2018	2008	2018				
0	Afghanistan	154.0	191.0	2040.0	23913.0	28950.0	10.79	55.52	NaN	27722.276	37171.921	1587.8	2190.2	-2.69	-2.75		
1	Albania	NaN	1003.0	3360.0	11588.0	15587.0	6175	115.74	3.0	3002.678	2882.740	10119.1	13601.3	-0.03	0.13		
2	Algeria	NaN	390.0	3322.0	27.0	278279.0	2473	91.83	2.0	34730.608	42228.408	10625.0	11479.5	-1.09	-0.79		
3	American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	57.492	55.465	NaN	NaN	0.98	1.21		
4	Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	83.862	77.006	NaN	NaN	1.34	1.43		
...		
233	Wallis and Futuna Islands	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	13.578	11.661	NaN	NaN	NaN	NaN		
234	Western Sahara	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	488.633	587.452	NaN	NaN	NaN	NaN		
235	Yemen	NaN	142.0	2154.0	95992.0	179768.0	10.38	54.52	NaN	21892.146	26498.683	NaN	NaN	-2.01	-3.03		
236	Zambia	1387.0	159.0	2002.0	10560.0	49487.0	13.70	60.53	NaN	12648.530	17551.708	2747.4	3521.5	0.47	0.14		
237	Zimbabwe	NaN	159.0	1908.0	3758.0	85837.0	11.59	44.15	2.0	12379.549	14438.802	1739.2	3130.0	-1.21	-0.71		

238 rows x 15 columns

Diagram illustrating variable groupings for a dataset:

- Disponibilités calorique et protéique** (Caloric and protein availability) points to `dispo_calories` and `dispo_proteines`.
- Millions d'habitants** (Millions of inhabitants) points to `population`.
- Croissance de la population de 2008 à 2018** (Population growth from 2008 to 2018) points to `population_croissance`.
- PIB par habitant** (GDP per capita) points to `ppa`.
- Ratio des importations de poulet par rapport aux exportations** (Ratio of chicken imports relative to exports) points to `poulet_import-export`.
- Ratio de protéines d'origine animale dans la disponibilité protéique** (Ratio of animal protein in protein availability) points to `ratio_proteines_animales`.
- PIB** (GDP) points to `pib`.
- Croissance du PIB de 2008 à 2018** (GDP growth from 2008 to 2018) points to `pib_croissance`.

	pays	dispo_calories	dispo_proteines	ratio_proteines_animales	population	population_croissance	pib	ppa	pib_croissance	poulet_import-export
0	Afghanistan	2040.0	55.52	0.194344	37.171921	1.340868	2190.2	0.000059	1.379393	155.279221
1	Albania	3360.0	115.74	0.533523	2.882740	0.960056	13601.3	0.004718	1.344122	NaN
2	Algeria	3322.0	91.83	0.269302	42.228408	1.215885	11479.5	0.000272	1.080424	NaN
3	American Samoa	NaN	NaN	NaN	0.055465	0.964743	NaN	NaN	NaN	NaN
4	Andorra	NaN	NaN	NaN	0.077006	0.918247	NaN	NaN	NaN	NaN
...
233	Wallis and Futuna Islands	NaN	NaN	NaN	0.011661	0.858816	NaN	NaN	NaN	NaN
234	Western Sahara	NaN	NaN	NaN	0.567402	1.210760	NaN	NaN	NaN	NaN
235	Yemen	2154.0	54.52	0.190389	28.498683	1.301777	NaN	NaN	NaN	NaN
236	Zambia	2002.0	60.53	0.226334	17.351708	1.350482	3521.5	0.000203	1.281757	7.599135
237	Zimbabwe	1908.0	44.15	0.262514	14.438802	1.166343	3130.0	0.000217	1.799678	NaN

238 rows × 10 columns

```
df.info()
```

→ Valeurs manquantes et infinies dans 132 pays

La majorité dans 'poulet_import_export'

→ Suppression des lignes concernées

→ 106 pays après nettoyage

```
RangeIndex: 238 entries, 0 to 237
```

index	pays	dispo_calories	dispo_proteines	ratio_proteines_animales	population	population_croissance	pib	pib_croissance	poulet_import-export	
0	0	Alghanistan	2040.0	55.52	0.194344	37.171921	1.340868	2190.2	1.379393	155.279221
1	7	Antigua and Barbuda	2446.0	80.19	0.645218	0.096289	1.127510	21115.8	0.914029	849.142857
2	8	Argentina	3307.0	106.77	0.646717	44.361150	1.106811	22745.9	0.984513	0.040777
3	9	Armenia	2997.0	94.35	0.480551	2.951745	1.015178	12715.0	1.214712	435.297297
4	11	Australia	3391.0	105.94	0.667359	24.898152	1.167158	49576.0	1.095894	0.145991
...
101	224	United Kingdom of Great Britain and Northern I...	3344.0	103.92	0.557737	67.141684	1.080402	46309.8	1.056045	1.279315
102	225	United Republic of Tanzania	2373.0	58.93	0.206007	56.313438	1.345475	2590.0	1.363517	43.307692
103	227	United States of America	3782.0	113.74	0.646035	327.096285	1.077797	61544.4	1.111030	0.019352
104	228	Uruguay	3202.0	86.50	0.534335	3.449285	1.032652	21590.8	1.362041	0.442524
105	236	Zambia	2002.0	69.53	0.226334	17.351708	1.350482	3521.5	1.281757	7.599135

```
106 rows x 10 columns
```

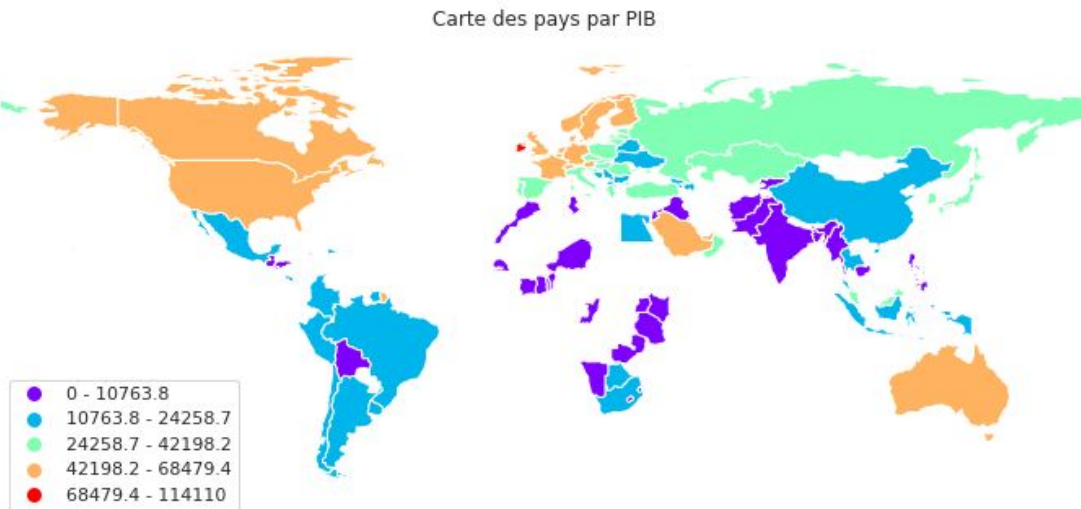
```
5    population_croissance    230 non-null    float64
6    pib                    185 non-null   float64
7    ppa                    185 non-null   float64
8    pib_croissance         184 non-null   float64
9    poulet_import-export   116 non-null   float64
dtypes: float64(9), object(1)
```


continent	pay	ion a3	neometrv
0	Oceania		
1	Africa	United Republic of Tanzi	
2	Africa	W. Sah	
3	North America	Can	
4	North America	United States of Ame	
...	...		
172	Europe	Se	
173	Europe	Montene	
174	Europe	Kos	
175	North America	Trinidad and Tob	
176	Africa	S. Su	

177 rows x 4 columns

Librairie **Geoplot**

- 177 pays disponibles
- Variable 'continent'



3	North America	United States of America	USA	MULTIPOLYGON (((112.294000 49.000000, -120.000000 49.0
---	---------------	--------------------------	-----	--

100 rows x 13 columns

pi_b_croissance poulet_import-export

1.310489 13.662252
1.363517 43.307692
1.061469 1.568819

83% de la population mondiale en 2018

```
['Afghanistan' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Australia'  
'Austria' 'Azerbaijan' 'Bahamas' 'Bangladesh' 'Barbados' 'Belarus'  
'Belgium' 'Benin' 'Bolivia' 'Bosnia and Herz.' 'Botswana' 'Brazil'  
'Bulgaria' 'Cambodia' 'Canada' 'Chile' 'China, Hong Kong SAR'  
'China, Macao SAR' 'China, mainland' 'Colombia' 'Congo' 'Costa Rica'  
'Croatia' 'Cyprus' 'Czechia' 'Côte d'Ivoire' 'Denmark'  
'Dominican Republic' 'Egypt' 'El Salvador' 'Estonia' 'Eswatini' 'Fiji'  
'Finland' 'France' 'Georgia' 'Germany' 'Ghana' 'Greece' 'Guatemala'  
'Honduras' 'Hungary' 'India' 'Indonesia' 'Iraq' 'Ireland' 'Italy'  
'Jamaica' 'Japan' 'Jordan' 'Kazakhstan' 'Kenya' 'Kuwait' 'Kyrgyzstan'  
'Latvia' 'Lebanon' 'Lesotho' 'Lithuania' 'Luxembourg' 'Malaysia' 'Malta'  
'Mexico' 'Morocco' 'Myanmar' 'Namibia' 'Netherlands' 'Niger'  
'North Macedonia' 'Norway' 'Oman' 'Pakistan' 'Peru' 'Philippines'  
'Poland' 'Portugal' 'Republic of Korea' 'Romania' 'Russia' 'Samoa'  
'Saudi Arabia' 'Senegal' 'Serbia' 'Slovakia' 'Slovenia' 'South Africa'  
'Spain' 'Suriname' 'Sweden' 'Switzerland' 'Thailand' 'Togo' 'Tunisia'  
'Turkey' 'Uganda' 'Ukraine' 'United Arab Emirates'  
'United Kingdom of Great Britain and Northern Ireland'  
'United Republic of Tanzania' 'United States of America' 'Uruguay'  
'Zambia']
```

Liste des 100 pays retenus

Classification hiérarchique

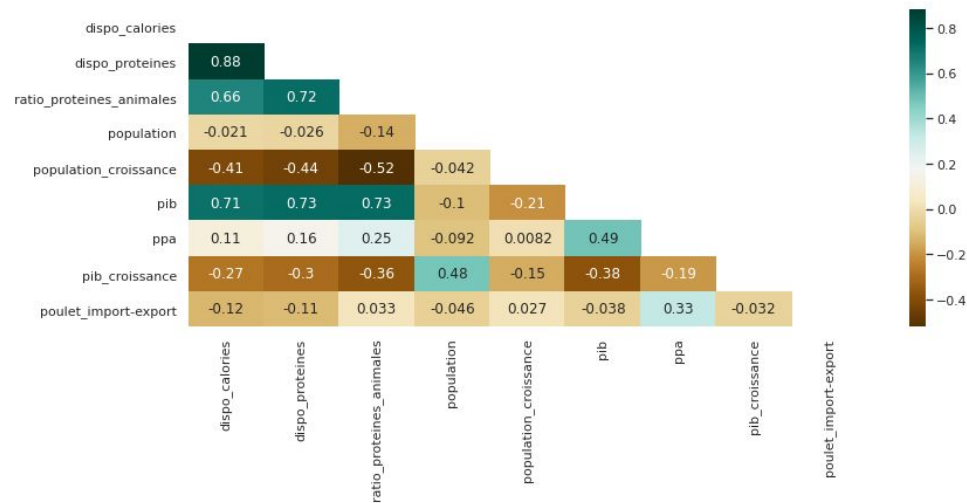
→ Peu de fortes corrélations

→ Dispo. en calories écartée

Très corrélée à la dispo. en protéines

Mêmes corrélations que la dispo. en protéines

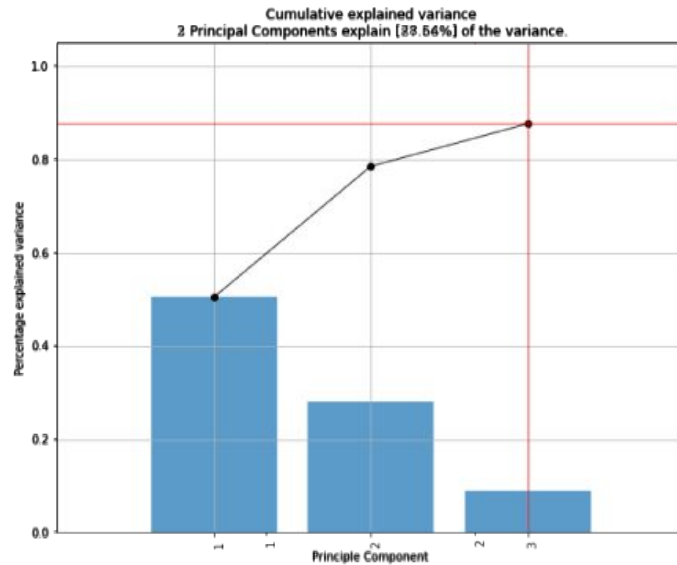
Matrice des corrélations



Variables centrées autour de 0

	ratio_proteines_animaux	population_croissance	pib_croissance	dispo_calories	dispo_proteines	population	pib	ppa	poulet_import-export
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	1.000000e+02	100.000000
mean	0.454784	1.132678	1.197033	0.227394	0.006225	0.005891	0.941274	1.907658e-07	0.030906
std	0.143988	0.152900	0.218673	0.192240	0.005115	0.021517	0.105064	3.599276e-07	0.115765
min	0.141283	0.871890	0.691271	0.030351	0.000952	0.000005	0.421799	6.761622e-10	0.000000
25%	0.347529	1.029482	1.051039	0.082764	0.002498	0.000214	0.953808	1.930648e-08	0.000030
50%	0.487820	1.112964	1.185242	0.158154	0.004274	0.000761	0.987013	7.515395e-08	0.000197
75%	0.568652	1.198523	1.302479	0.280879	0.007577	0.003388	0.996512	1.837584e-07	0.002643
max	0.680987	1.755557	2.056280	0.905994	0.029119	0.190422	0.999539	2.266971e-06	0.778994

Variables non centrées autour de 0



Variance expliquée

ACP

- Réduction des variables à x dimensions
- Utile pour visualiser ou modéliser
- Sur les variables normalisées

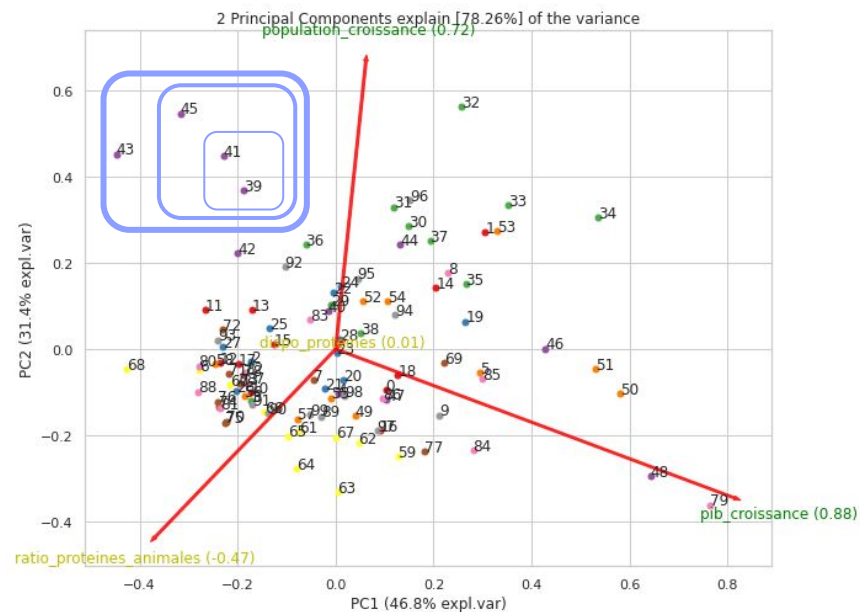
Variables écartées

- Variables qualitatives
- Dispo. calories

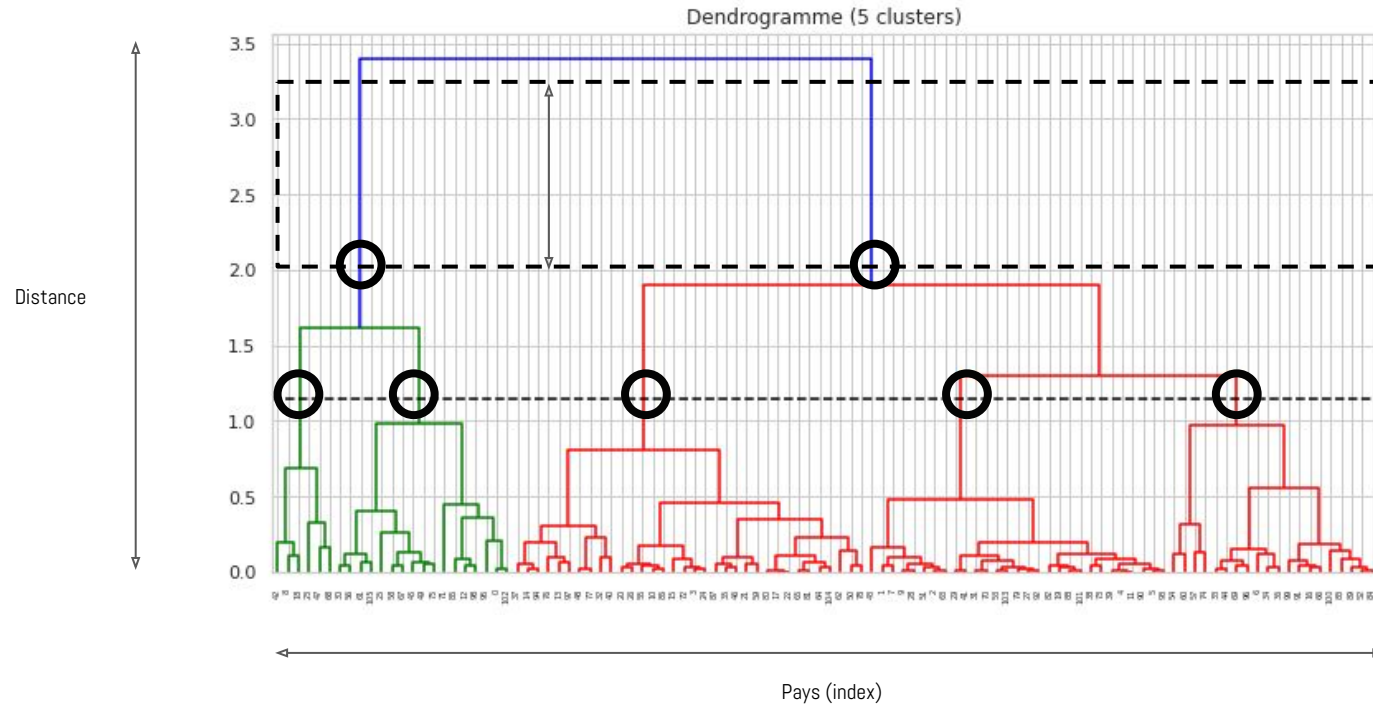
Variance expliquée

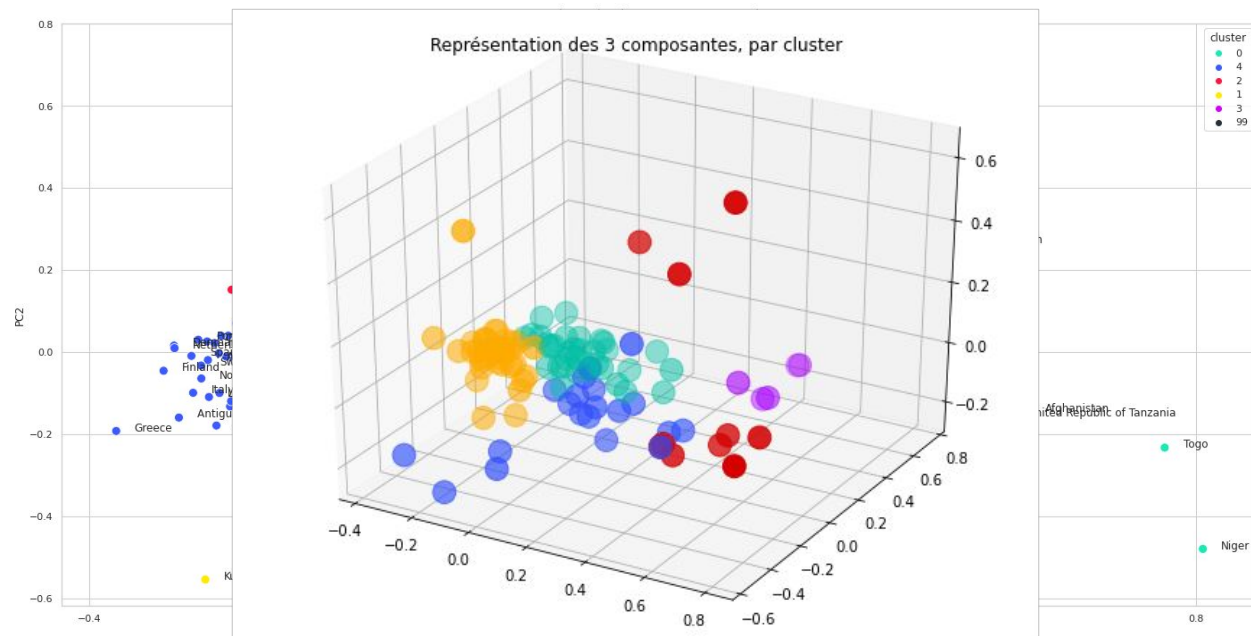
- 2 composantes : 78%
- 3 composantes : 87%

Composante 2
(variance expliquée : 31%)



Composante 1
(variance expliquée : 46%)





Variable testée : disponibilité protéique

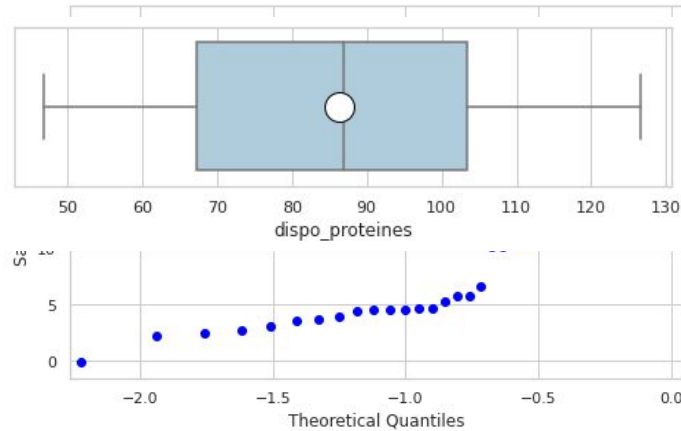


Diagramme quantile-quantile de 'dispo_proteines'
(après boxcox)

Normalité des résidus

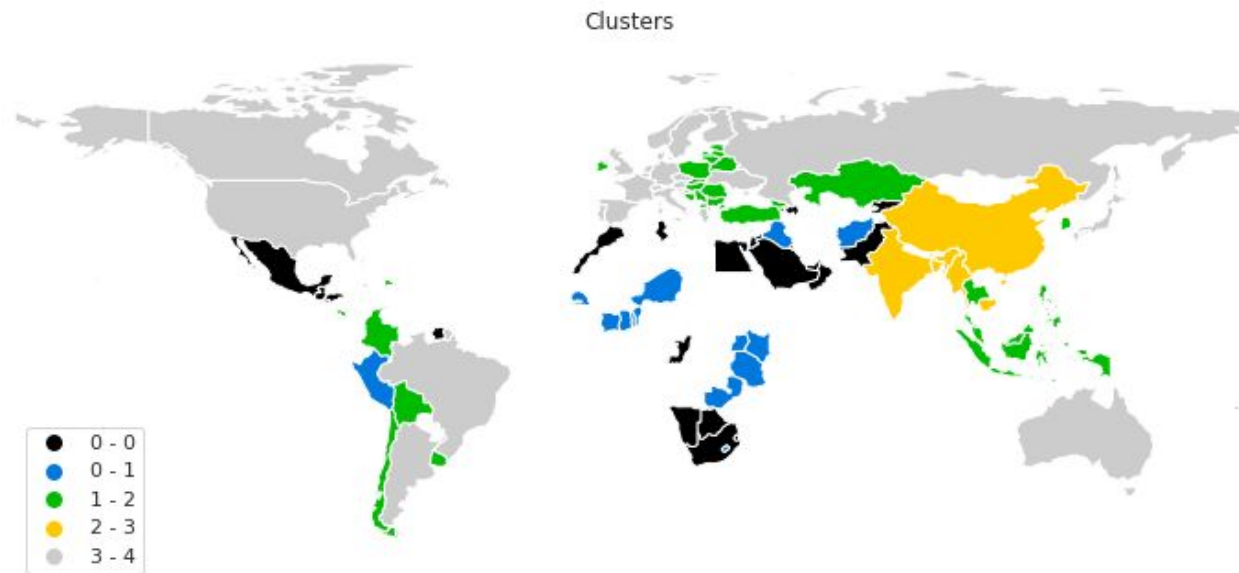
- Shapiro (avec boxcox)
- p-value : 100%
- H_0 : distribution normale

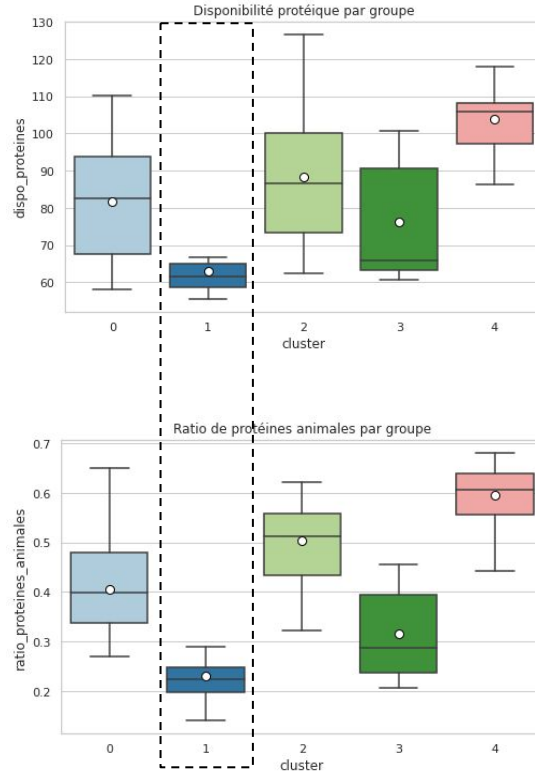
Test de comparaison entre 2 clusters

- Kolmogorov-Smirnov
- Entre les clusters 0 et 3
- p-value : 51%
- H_0 : les 2 échantillons viennent d'une population ayant la même distribution

Analyse des groupes

4



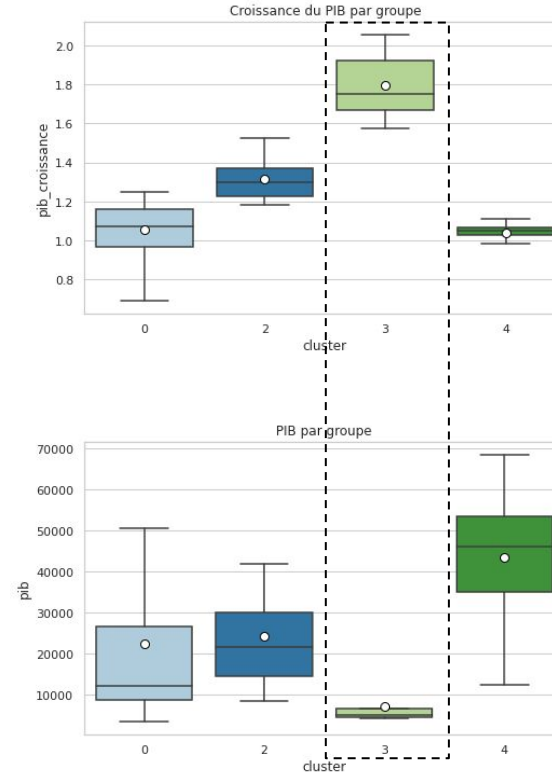


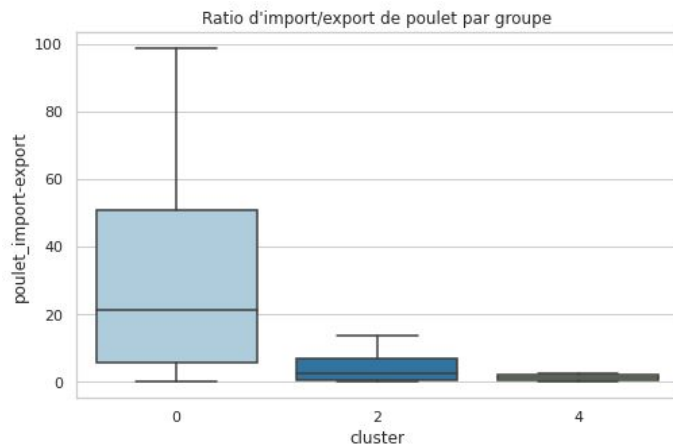
Groupe 1

- La plus faible disponibilité protéique
- Le plus faible ratio de protéines animales
- 14 pays écartés
United Republic of Tanzania, Kenya, Lesotho, Peru, Senegal, Benin, Niger, Togo, Ghana, Côte d'Ivoire, Zambia, Iraq, Afghanistan, Uganda

Groupe 3

- La plus forte croissance du PIB
- Les plus faibles PIB et PPA
- 5 pays écartés
Cambodge, Myanmar, Inde, Bangladesh, Chine





Groupe 4

→ Le plus faible ratio d'import / export

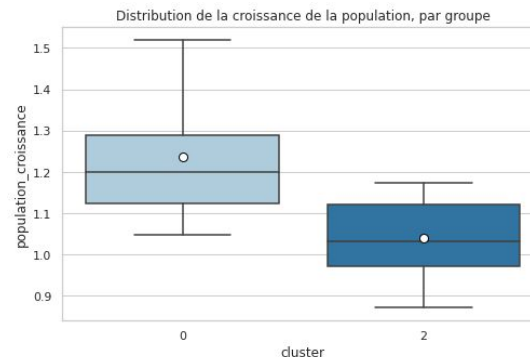
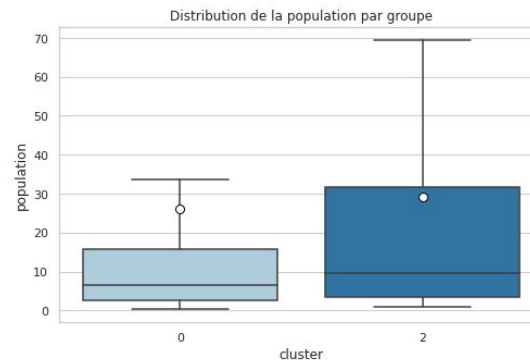
→ 26 pays écartés

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

Groupes 2

- Moyenne de population proche du groupe 0
- Croissance de la population plus faible
- 29 pays écartés

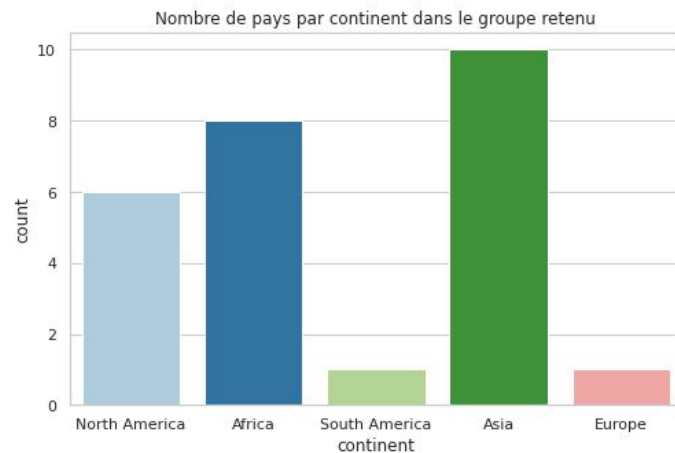
Fiji, Kazakhstan, Indonesia, Chile, Dominican Republic, Uruguay, Bolivia (Plurinational State of), Colombia, Costa Rica, Thailand, Republic of Korea, Armenia, Belarus, Poland, Hungary, Romania, Lithuania, Latvia, Estonia, Bulgaria, Turkey, Ireland, Georgia, Philippines, Malaysia, Slovakia, Bosnia and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



5 groupes de pays →	0	1	2	3	4
Consommation de protéines animales	+	-	+	+	++
PIB	+		+	-	++
Importations de poulet	++		+		-
Croissance de la population	++		+		

Groupe 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



Classement du cluster 1 par ratio de protéines animales



Cluster 1 par croissance de la population

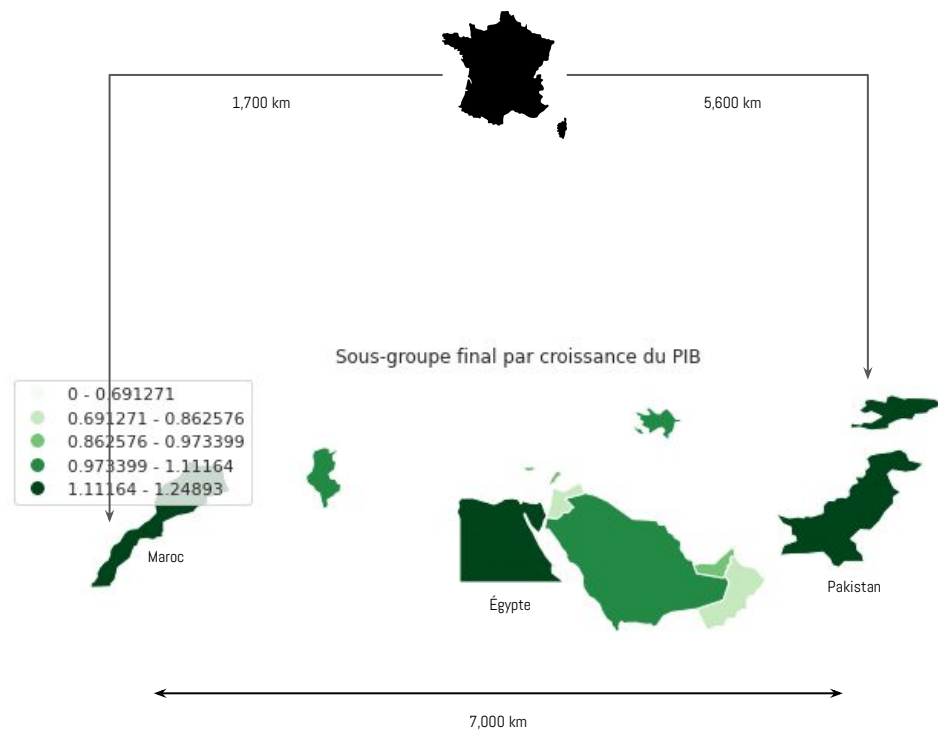


Cluster 1 par croissance du PIB



Cluster 1 par ratio d'import de poulet





Shortlist : 13 pays

- Arabie saoudite
- Azerbaïjan
- Chypre
- Égypte
- Émirats arabes unis
- Jordanie
- Koweït
- Kirgizstan
- Liban
- Maroc
- Pakistan
- Sultanat d'Oman
- Tunisie

Résumé

- Analyse sur 83% de la population mondiale
- ACP à 78% de variance expliquée
- 5 clusters par classification hiérarchique

Recommandations

- Cibler 13 pays du Maroc au Kirgizstan
- Former une équipe arabophone
- Créer une sous-marque adaptée