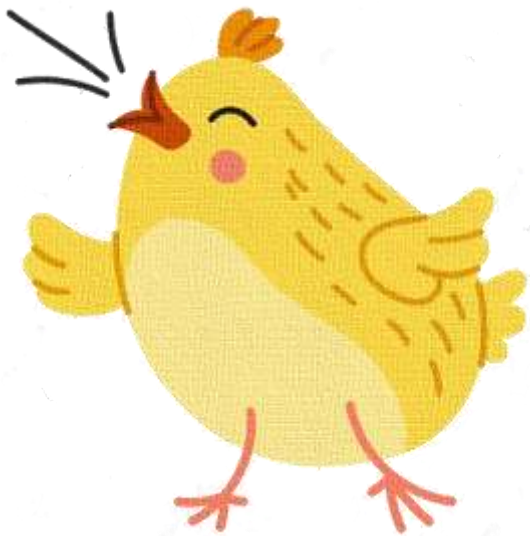


# Etude de marché sur l'exportation de poulet dans le monde

Par OpenClassRooms

## La poule qui chante



### Contexte du projet

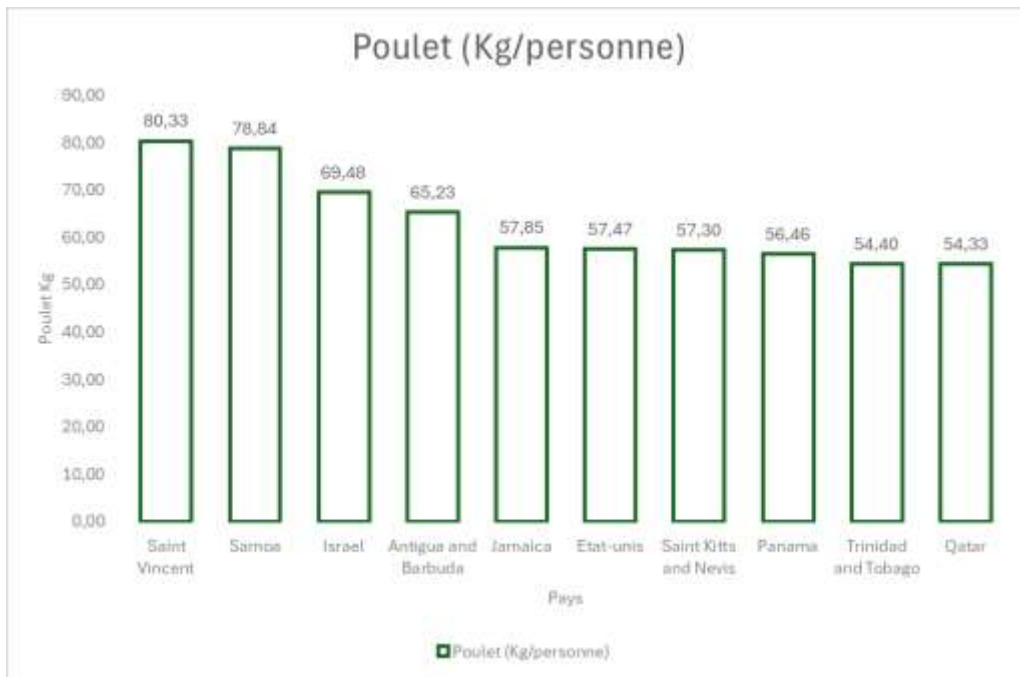
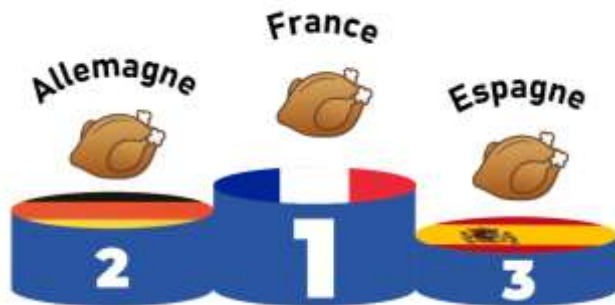
- ❖ Activité : Élevage et la vente de poulets sous le label “Poulet Agriculture Biologique”
- ❖ Souhaite se développer à l’international
- ❖ A employer un data analyst pour répondre aux besoins

### Identifier les pays pour exporter les poulets:

- ❖ Collecte de données pour affiner les résultats
- ❖ Préparation et nettoyage des données
- ❖ Réalisation d’un clustering hiérarchique / Kmeans
- ❖ Analyse d’un ACP pour les similitudes entre variables
- ❖ Recommandations des pays / Regions

# Analyse de la consommation de poulet

**La France : 1<sup>er</sup> pays consommateur de volaille de l'Union européenne**



# Analyse de la consommation de poulet

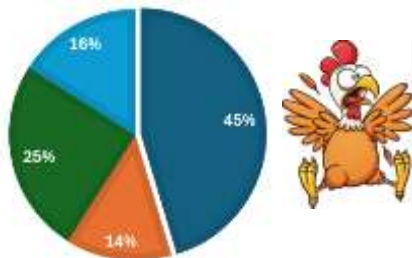
■ Poulet (Kg/personne)

■ Bœuf (Kg/personne)

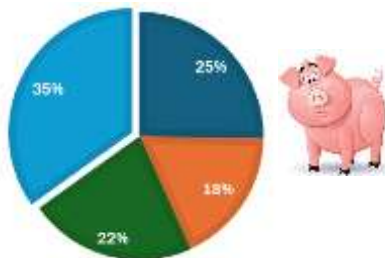
■ Poissons (Kg/personne)

■ Autres viandes (Kg/personne)

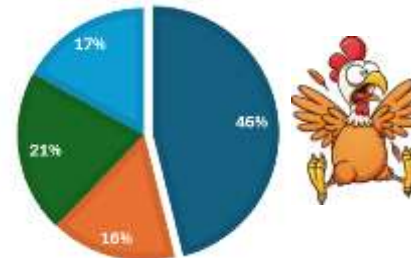
MÉDITERRANÉE ORIENTALE



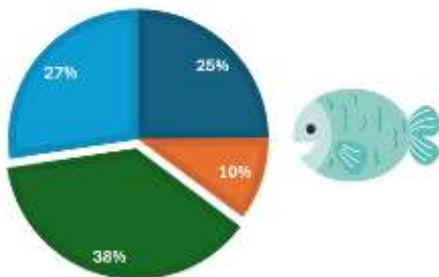
EUROPE



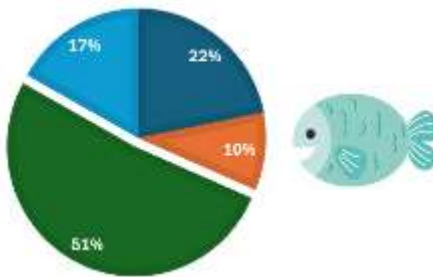
AMERIQUE



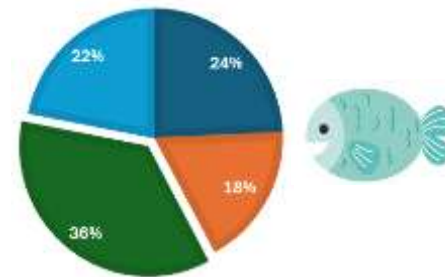
PACIFIQUE OCCIDENTAL



ASIE DU SUD-EST

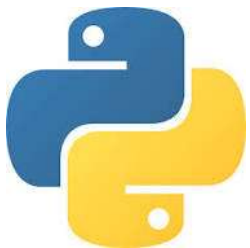


AFRIQUE



# Méthodologie de l'analyse

## Outils utilisés



Langage de programmation, conçu pour les analyses de données



Jupyter est un environnement de développement interactif basé sur le web pour les notebooks, le code et les données.

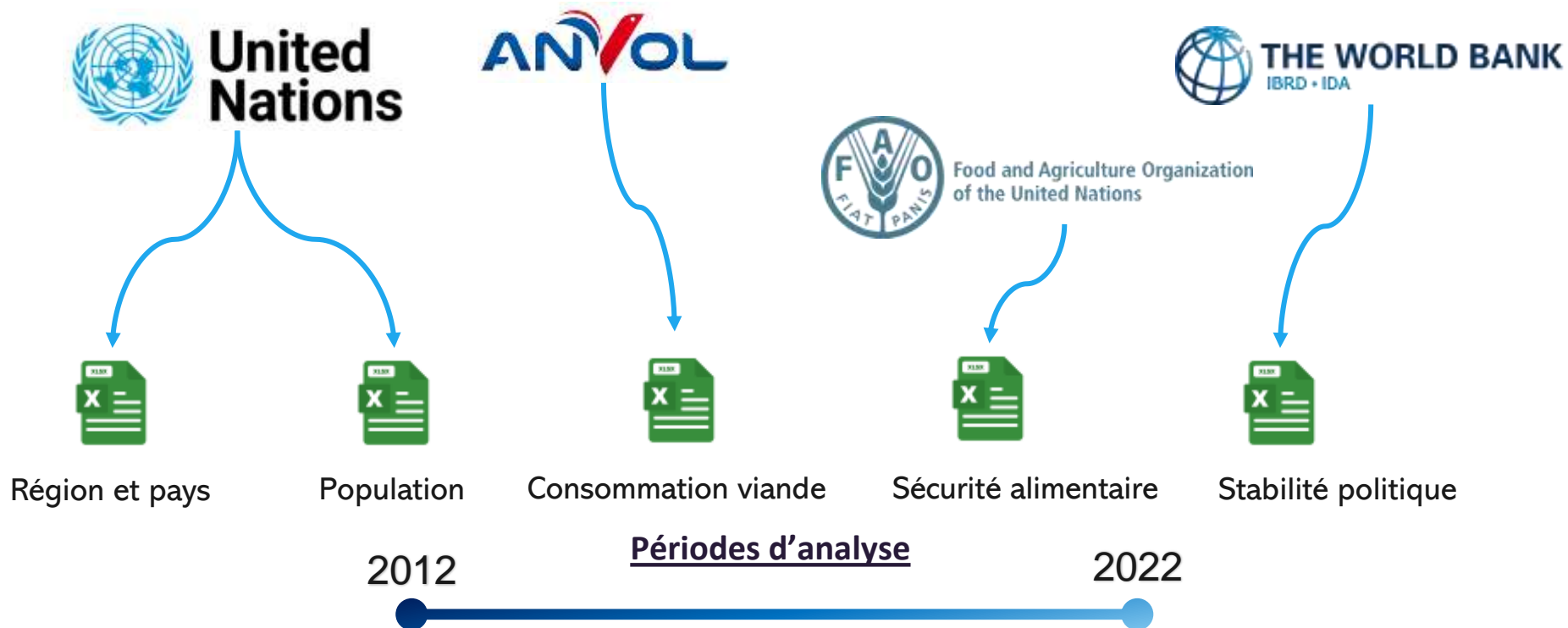
## Ressources



Poulet\_qui\_chante.csv

- Région
- Population
- Consommation\_viande
- Import\_poulet
- Production\_Poulet
- Stabilité\_politique
- Sécurité\_alimentaire
- PIB\_Pays

# Collecter les données & Sources utilisées



# Préparation & Nettoyage des données

Pour chaque dataframe, je vérifie:

- ❖ Les informations (colonne, ligne, type de donnée)
- ❖ Les valeurs manquantes
- ❖ Une statistique descriptive (mean, min, max)
- ❖ Les valeurs aberrantes

Information sur df\_pib

```
df_pib.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1351 entries, 0 to 1350
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Country                1351 non-null  object  
1   Year                   1351 non-null  int64   
2   PIB_par_habitant ($)   1319 non-null  float64  
dtypes: float64(1), int64(1), object(1)
memory usage: 31.8+ KB
```

Vérifier les valeurs manquantes par pays

Country	Country	Year	PIB_par_habitant (\$)
Afghanistan	0	0	1
Bhutan	0	0	1
Cuba	0	0	2
Democratic People's Republic of Korea	0	0	6
Eritrea	0	0	4
Lebanon	0	0	1
Liechtenstein	0	0	1
Palau	0	0	1
San Marino	0	0	1
South Sudan	0	0	6
Syrian Arab Republic	0	0	1
Tonga	0	0	1
Venezuela (Bolivarian Republic of)	0	0	6

Statistiques descriptives

	PIB_par_habitant (\$)
count	1351.0
mean	16306.0
std	26633.0
min	217.0
25%	2065.0
50%	6095.0
75%	18225.0
max	240862.0

# Consolidations des données

## Jonction des fichiers à l'aide de la colonne Country et Year

```
# Fusionner df_region et df_population
df_merge = pd.merge(df_region, df_population, on='Country', how='inner')

# Fusionner le résultat avec df_Conso_mation_Viande
df_merge = pd.merge(df_merge, df_Conso_mation_Viande, on=['Country', 'Year'], how='inner')
# Fusionner le résultat avec df_Import_Export
df_merge = pd.merge(df_merge, df_Import_Export, on=['Country', 'Year'], how='inner')
# Fusionner le résultat avec df_Production_Poulet
df_merge = pd.merge(df_merge, df_Production_Poulet, on=['Country', 'Year'], how='inner')
```

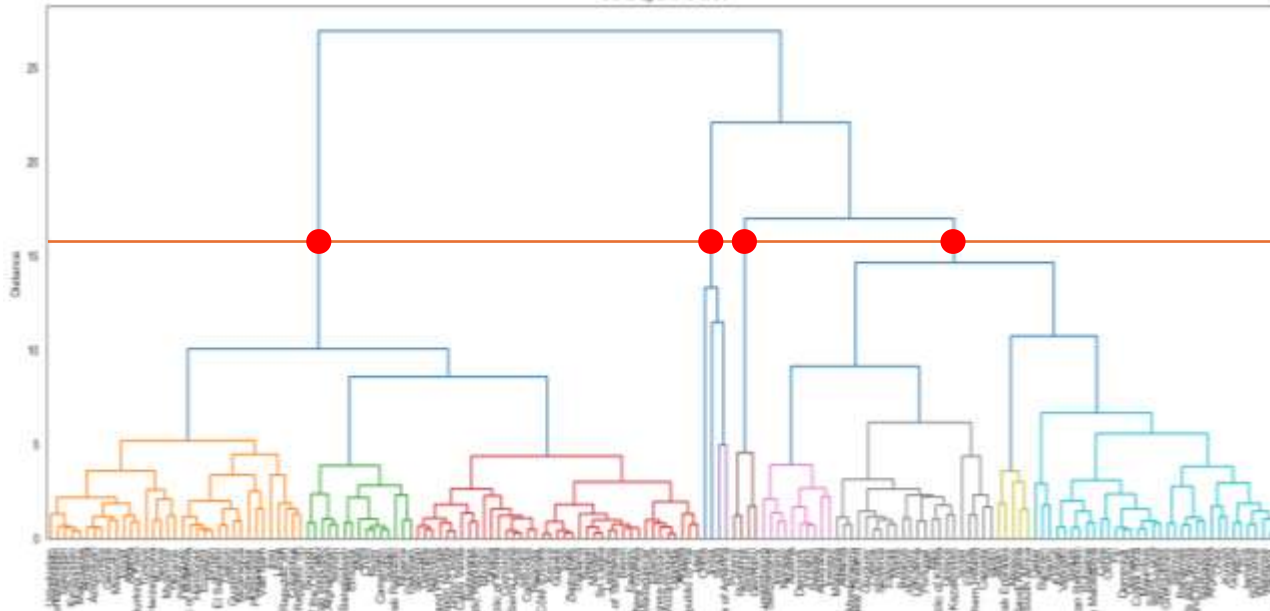
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 671 entries, 0 to 670
Data columns (total 22 columns):
```

- On a 168 pays
- 671 lignes réparties sur 4 années
- On a choisi l'année 2019 qui est le plus adapté à notre analyse



# Clustering hiérarchique : Dendrogramme CAH

Dendrogramme CAH



On coupe le dendrogramme en 4 clusters :

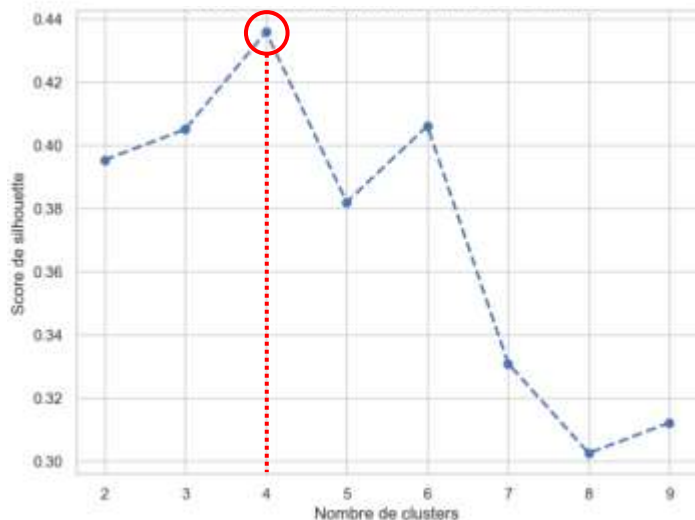
Cluster	Pays
1	89
2	4
3	4
4	71

*On vérifiera la médiane des variables par clusters en comparant la méthode CAH et Kmeans (slide 8)*

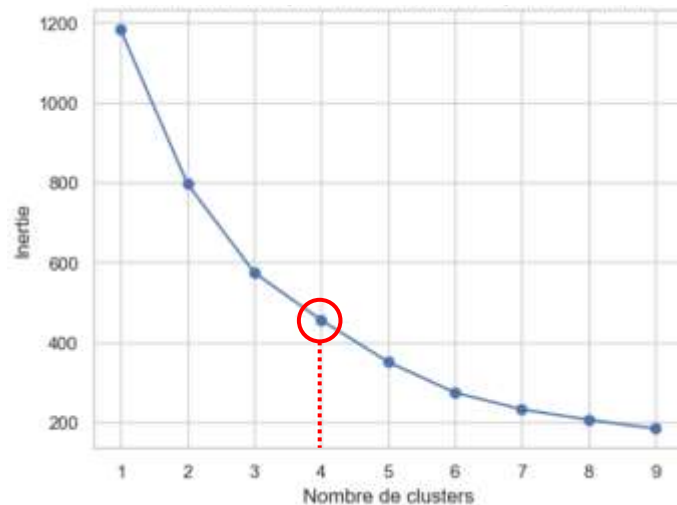
# Clustering avec méthode Kmeans

Nous allons utiliser deux méthodes pour connaître le nombre de cluster optimal

Score de silhouette



Méthode du coude



Le nombre de cluster optimal est de 4

Nombre de valeurs uniques par cluster :

```
cluster
1      89
4      71
3       4
2       4
```

# Clustering avec méthode Kmeans

## Médiane des variables par cluster avec kmeans

	Population	Consommation	Importations_têtes	Importations_tonnes	Production
cluster					
1	16126866.0	7.0	607.0	7462.0	67234.0
2	855721002.0	31.0	948.0	32769.0	18041758.0
3	28668562.0	20.0	129388.0	155062.0	1025486.0
4	4441100.0	33.0	279.0	18464.0	71651.0

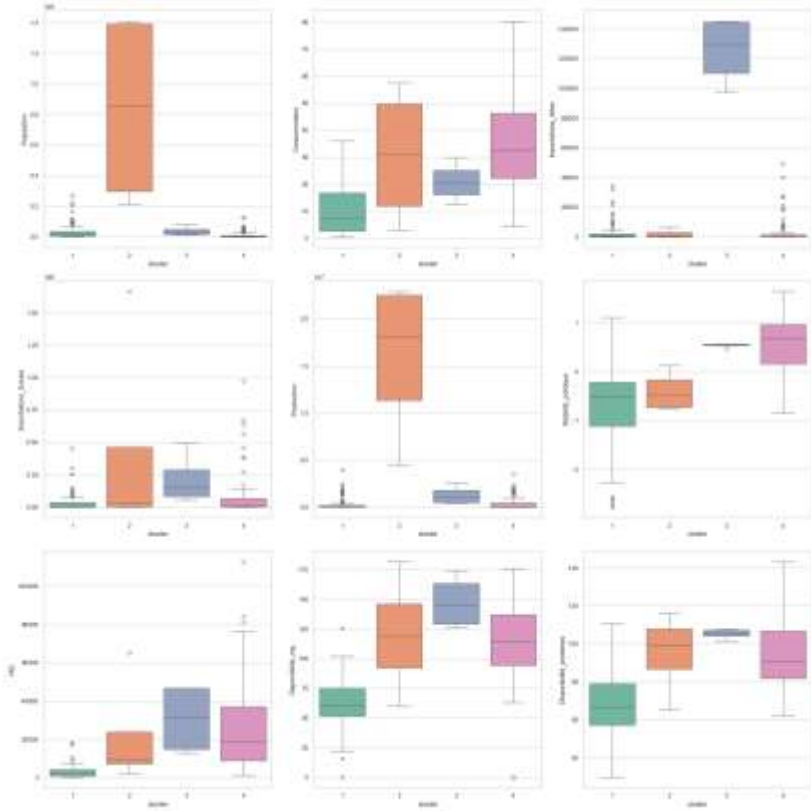
## Médiane des variables par cluster avec le CAH

	Population	Consommation	Importations_têtes	Importations_tonnes	Production
Cluster_CAH					
1	16126866.0	7.229999	607.0	7462.220	67234.0
2	855721001.5	30.920000	948.5	32769.035	18041757.5
3	28668561.5	20.419998	129387.5	155062.315	1025486.0
4	4441100.0	32.897423	279.0	18463.560	71651.0

CAH = Kmeans  
Avec ces données!

Ce qui signifie que les  
données sont bien  
structurées

# Analyse des clusters



## CLUSTER 1 :

- Stabilité politique
- Disponibilité en protéines
- Disponibilité en matière grasse

Faible disponibilité en matière grasse, PIB bas, consommation modérée de poulet. Besoin potentiel d'aide alimentaire ciblée.

## CLUSTER 2 :

- Population
- Production de poulet (tonnes)
- Consommation de poulet par personne
- Disponibilité en matière grasse
- Importations de poulet (tonnes)

Population importante, PIB moyen, forte production locale de poulet avec faibles importations. Marché du poulet déjà développé.

## CLUSTER 3 :

- Importations de poulet (têtes)
- Disponibilité en matière grasse
- PIB

Population modeste, importations élevées de poulet, production locale limitée. Opportunités d'intervention économique spécifique.

## CLUSTER 4 :

- Consommation de poulet par personne
- Stabilité politique
- Disponibilité en matière grasse
- Disponibilité en protéines


Faible production locale mais forte consommation de poulet, stabilité politique solide. Principalement des pays européens.

# Etape réalisation d'un ACP

## Standardiser les Données

- Mettre les données à la même échelle avec les variances équitables

ID	Car	Pop.
1	128	12456
2	234	36423
3	5	248
4	112	1958



ID	Car	Pop.
1	0,10	-0,02
2	1,41	1,64
3	-1,41	-0,87
4	-0,10	-0,75

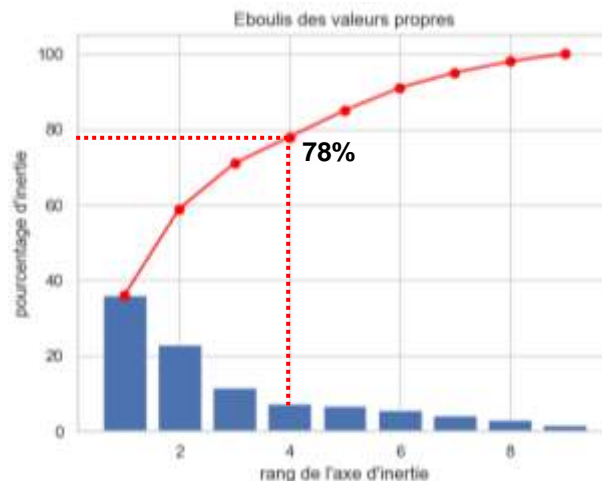
*Les moyennes sont à 0 et l'écart-type sont à 1!*

## Effectuer l'ACP

- Permet de réduire le nombre de variables
- Identifier et visualiser des clusters dans des jeux de données complexes

## Eboulis des valeurs propres

- Permet de connaître le nombre de composant principales à utiliser

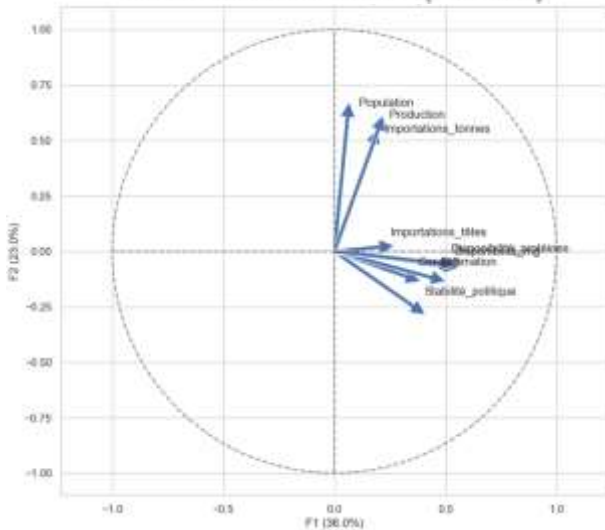


# Heatmap entre les variables et les composantes



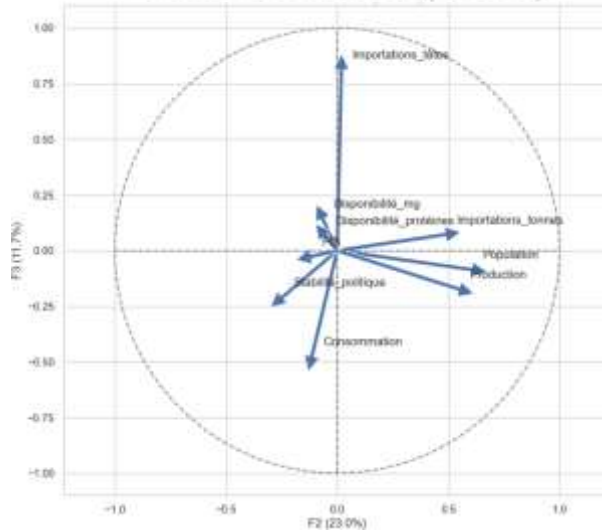
# Cercle des corrélations : Interprétation

Cercle des corrélations (F1 et F2)



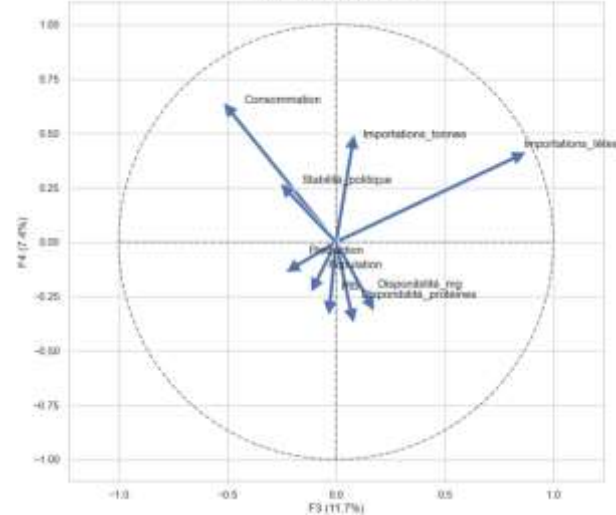
**F1 :** Pays avec une consommation élevée, une stabilité politique, un PIB, et une disponibilité en nutriments et protéines

Cercle des corrélations (F2 et F3)



**F3 :** Importation\_1000 têtes et consommation, stabilité politique

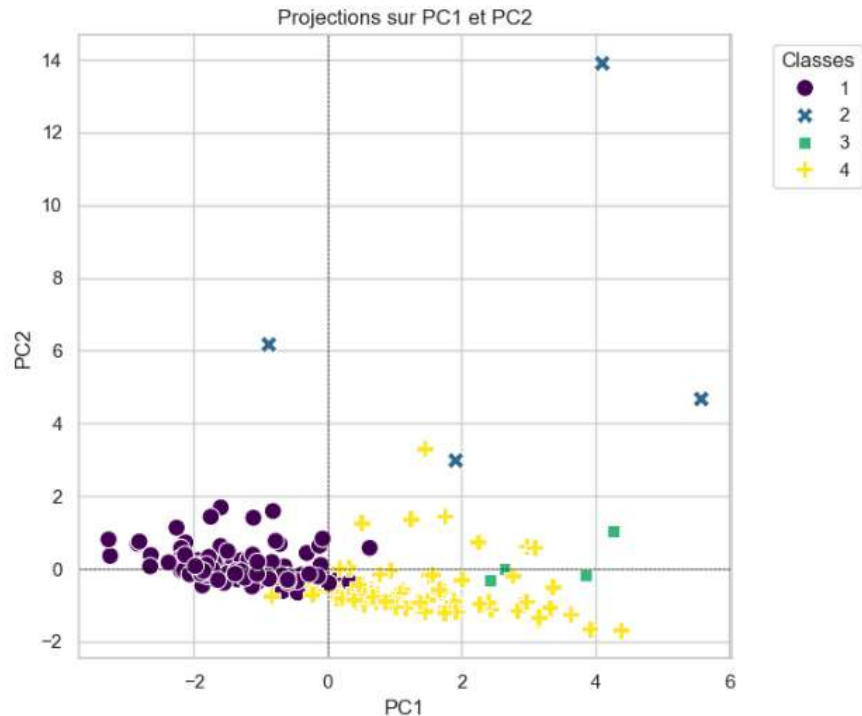
Cercle des corrélations (F3 et F4)



**F4 :** Consommation, Importation tonne

**F2 :** Pays avec une grande population, de fortes importations en tonnes, et une production élevée (tonnes)

# Projection des individus : Interprétation



**Cluster 1** : Corrélation négative avec PC1, indiquant des pays moins développés économiquement, mais une corrélation positive avec PC2, suggérant une tendance à importer et produire du poulet.

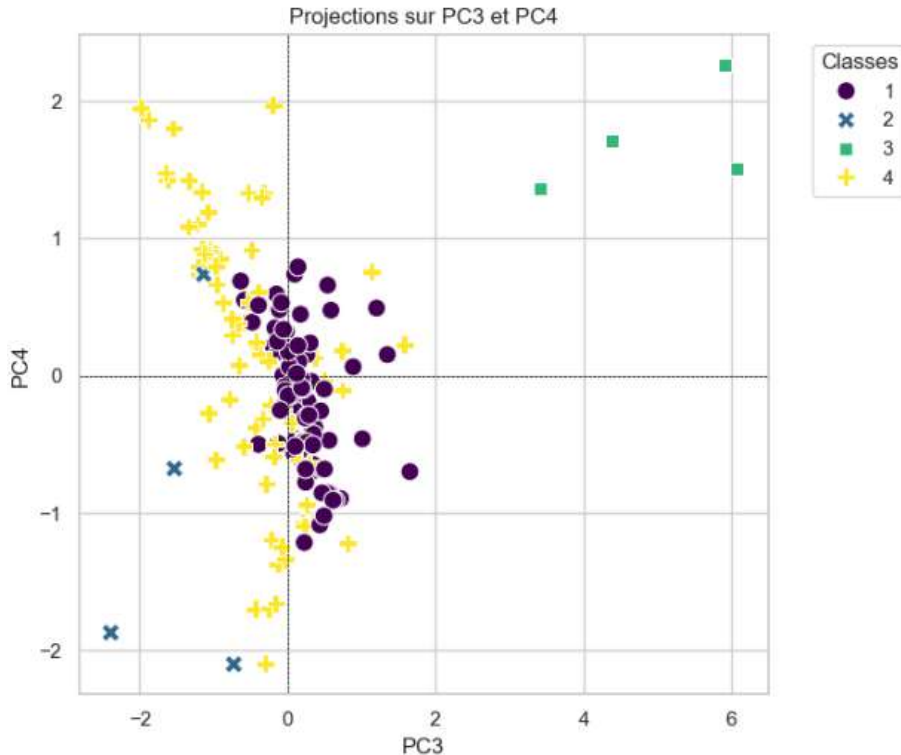
**Cluster 2** : Il y a uniquement quatre pays. Ils sont fortement corrélés avec PC1 et PC2. pays avec une grande population et une production élevée

**Cluster 3** : uniquement 4 pays, il est corrélé positivement avec PC1 mais peu avec PC2. pays avec une production et une population moyenne mais des importations en tête élevées

**Cluster 4** : corrélation positive avec PC1 mais négative avec PC2 donc pays développé mais importe peu de poulet malgré une consommation élevé



# Projection des individus : Interprétation



**Cluster 1** : Peu de corrélation avec ces composantes.

**Cluster 2** : Il y a uniquement quatre pays. Ils ne sont pas corrélés avec PC1 et PC2. Ce sont des pays qui importe peu de poulet par rapport au nombre de personne et leur consommation n'est pas conséquente.

**Cluster 3** : uniquement 4 pays, il est corrélé positivement avec PC1 et PC2. On peut donc conclure que ces pays importent et consomme beaucoup

**Cluster 4** : Des pays extrêmement opposer, ils sont généralement corrélés négativement avec PC1, indiquant des valeurs élevées en Consommation, Importations\_tonnes, et Disponibilité\_protéines



## Cluster 3

Intéressant pour la vente de poulet car ces pays importent déjà beaucoup et peuvent être ouverts à davantage d'importations



## Cluster 4

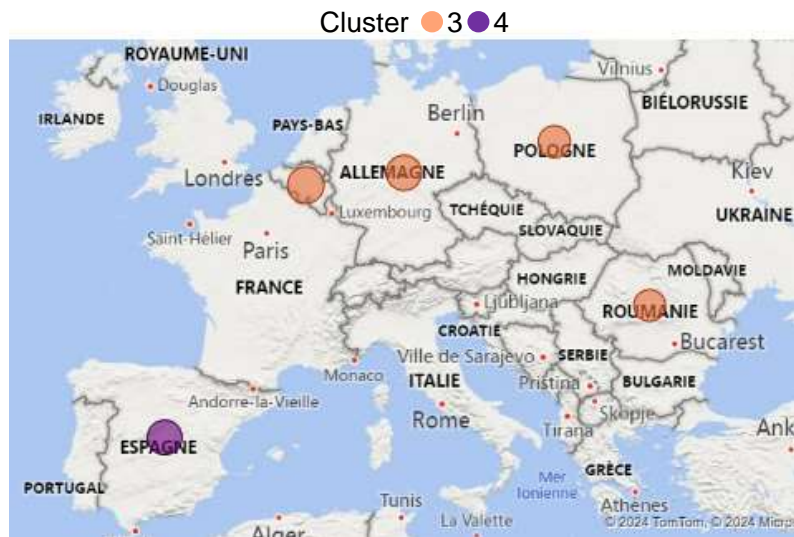
Très intéressant car ces pays montrent une demande élevée, un pouvoir d'achat suffisant, et une ouverture aux importations

# Choix des pays

J'ai utilisé une carte pour visualiser la répartition des pays selon leurs clusters

J'ai filtré les données pour ne conserver que les pays avec des niveaux élevés d'importation et de consommation de poulet.

J'ai identifié principalement des pays européens.



J'ai filtré les données pour ne conserver que les pays avec les disponibilités de matière grasse très faible mais avec une bonne consommation de poulet.

J'ai identifié des pays de l'Asie du Sud-Ouest

