

## Projet 9:

Produisez une étude de marché avec **R** ou  
**Python**



# La poule qui chante

# Introduction

Dans le but de développer l'entreprise vers l'international on va réaliser une étude du marché et grouper les pays pour bien définir les pays à viser dans la nouvelle stratégie.

## Mission

Développement d'un outil d'aide à la décision de quels pays viser par la nouvelle politique.

# Démarches

On va tout d'abord commencer par la préparation, nettoyage et analyse exploratoire de nos données (Chargement des données, gestion des valeurs manquantes, des doublons, etc..., exploration des variables, identification des variables pertinentes)  
Puis la normalisation des données et leur centrage et réduction pour à la fin réaliser des algorithmes de regroupement à l'aide de l'outil Python:

## Etape 1

### Analyse en composantes principales (ACP):

- Présentation de l'ACP comme méthode de réduction de dimension
- Utilisation de l'ACP pour visualiser les résultats du clustering
- Interprétation des groupes, des liens entre les variables et les individus

## Etape 2

### Méthode des k-Means

- Description de la méthode des k-means et de son utilisation pour le clustering
- Application des k-Means sur les données pour affiner l'analyse
- Analyse des centroïdes des classes pour comprendre les caractéristiques de chaque

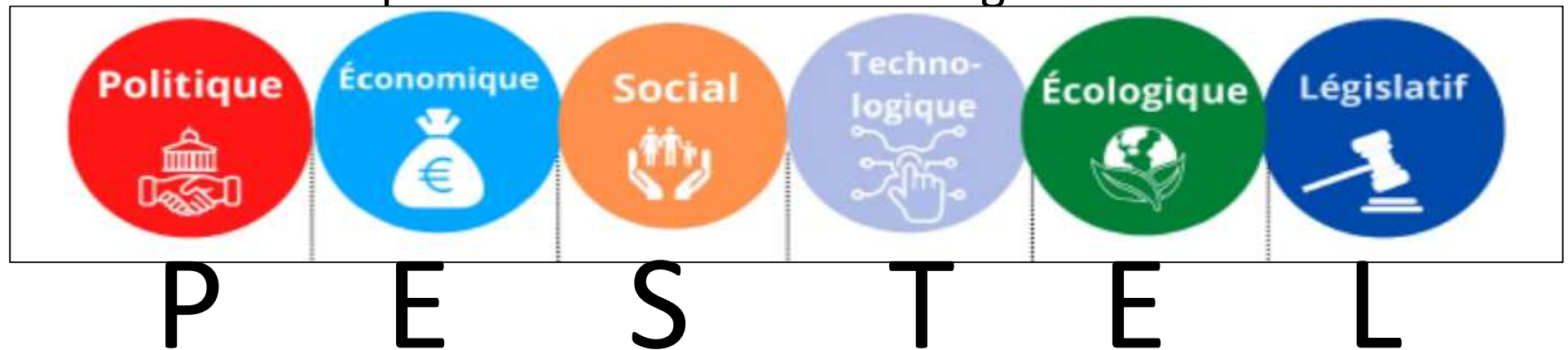
## Etape 3

### Classification ascendante hiérarchique (CAH)

- Explication de la méthode CAH et de son fonctionnement
- Application de la CAH sur les données pour créer des groupes de pays similaires
- Analyse des résultats
- Comparaison des résultats obtenus avec ceux de la CAH

# ANALYSE PESTEL

On utilise la méthode PESTEL pour analyser la stratégie et pour comprendre les facteurs susceptibles d'influencer la décision. PESTEL signifie Politique, Économique, Social, Technologique, Environnemental et Légal. Elle est utilisée pour identifier les opportunités et les menaces dans l'environnement commercial, aidant les organisations à prendre des décisions éclairées et à s'adapter aux circonstances changeantes.



# Les jeux de données

## Disponibilité\_alimentaire

- L'analyse de la disponibilité alimentaire permet d'évaluer la capacité du pays cible à fournir les matières premières nécessaires à la production de poulet, comme les céréales et les protéines animales.

0	Zone	172 non-null
1	Disponibilité intérieure	170 non-null
2	Exportations - Quantité	135 non-null
3	Importations - Quantité	170 non-null
4	Nourriture	170 non-null
5	Pertes	67 non-null
6	Production	168 non-null
7	Variation de stock	169 non-null

## PIB

- L'analyse du Produit Intérieur Brut (PIB) des pays cibles est un facteur clé car il reflète la taille de l'économie d'un pays, et il est un indicateur de la capacité des consommateurs à dépenser.

0	Pays	237 non-null
1	PIB_Habitant	209 non-null

## Stabilité politique

- évaluer des risques politiques potentiels, de privilégier les pays stables, compréhension de l'environnement réglementaire et de saisir les opportunités commerciales.

0	Pays	198 non-null
1	Stabilité_pol	198 non-null

## Population 2000<2018

- L'analyse de la population entre 2000 et 2018 permet de comprendre l'évolution démographique dans les pays cibles, ce qui a un impact sur la demande de la viande de volaille.

0	Pays	236 non-null
1	% Popu 2007_2017	229 non-null
2	Population	236 non-null

## Revenus

- L'analyse des classes de revenus permet de hiérarchiser les pays en fonction de leur niveau de revenu par habitant, ce qui peut influencer la consommation à acheter vos produits avicoles.

0	Pays	265 non-null
1	Groupe_revenu	264 non-null

## Taux total d'impôts et de cotisations

- Le taux d'imposition peut avoir un impact sur les coûts de production, ce qui est pertinent pour l'industrie de la volaille.

0	Pays	266 non-null
1	TTCR	236 non-null

## Investissements étrangers directs

- Les investissements étrangers dans le secteur agricole ou agroalimentaire peuvent influencer la production et la distribution de viande de volaille.

0	Pays	264 non-null
1	IED	247 non-null

## Dépenses liées au tourisme international

- Le tourisme peut stimuler la demande de produits alimentaires, y compris de la volaille, dans les destinations touristiques.

0	Pays	266 non-null
1	ITE	204 non-null

## Inflation, prix à la consommation

- L'inflation peut avoir un impact sur les coûts de production et les prix des aliments, ce qui peut affecter la demande de viande de volaille.

0	Pays	266 non-null
1	ICP	229 non-null

## Formation brute de capital fixe

- Cela peut refléter les investissements dans l'infrastructure agricole et l'industrie de la volaille

0	Pays	266 non-null
1	GCF	225 non-null

## Indice de profondeur de l'information sur le crédit

- Cette variable pourrait indiquer la facilité d'accès au crédit pour le secteur agricole, ce qui est important pour le développement de l'industrie de la volaille.

0	Pays	266 non-null
1	DCII	237 non-null

# D'autres Variables qui auront être utiles

On a essayé plusieurs tentatives comme par exemples travailler avec 20 variables, éliminer des individus des data sets et éliminer une variables puis une autre mais globalement on arrive au même résultat parce que plusieurs variables ont tendance a ne contenir que des valeurs nulles après les jointures.

## Taux total d'impôts et de cotisations

- Ce facteur peut influencer la rentabilité des entreprises. Un taux élevé d'impôts et de cotisations peut réduire les marges bénéficiaires des entreprises.

DCII 48 non-null

## Dépenses liées au tourisme international

- Les dépenses des touristes internationaux peuvent avoir un impact indirect. Les habitudes alimentaires des touristes peuvent affecter la demande de produits alimentaires locaux, y compris la volaille. Par conséquent, si la région est une destination touristique, cela peut influencer la demande de volaille.

GCF 45 non-null

## Inflation, prix à la consommation

- L'inflation peut affecter les coûts de production et les prix des produits alimentaires. Une inflation élevée peut augmenter les coûts de production, ce qui peut être répercuté sur les prix pour les consommateurs.

TTCR 48 non-null

## Indice de profondeur de l'information sur le crédit

- Cet indice peut refléter la facilité d'accès au financement pour les entreprises. Un accès plus facile au crédit peut favoriser l'expansion des entreprises et la modernisation des installations, ce qui peut influencer la production et la qualité des produits.

ITE 43 non-null

## Formation brute de capital fixe

- Cet indicateur mesure les investissements dans les infrastructures et les équipements. Il peut être lié à l'amélioration des conditions d'élevage, de la chaîne d'approvisionnement et de la qualité du produit

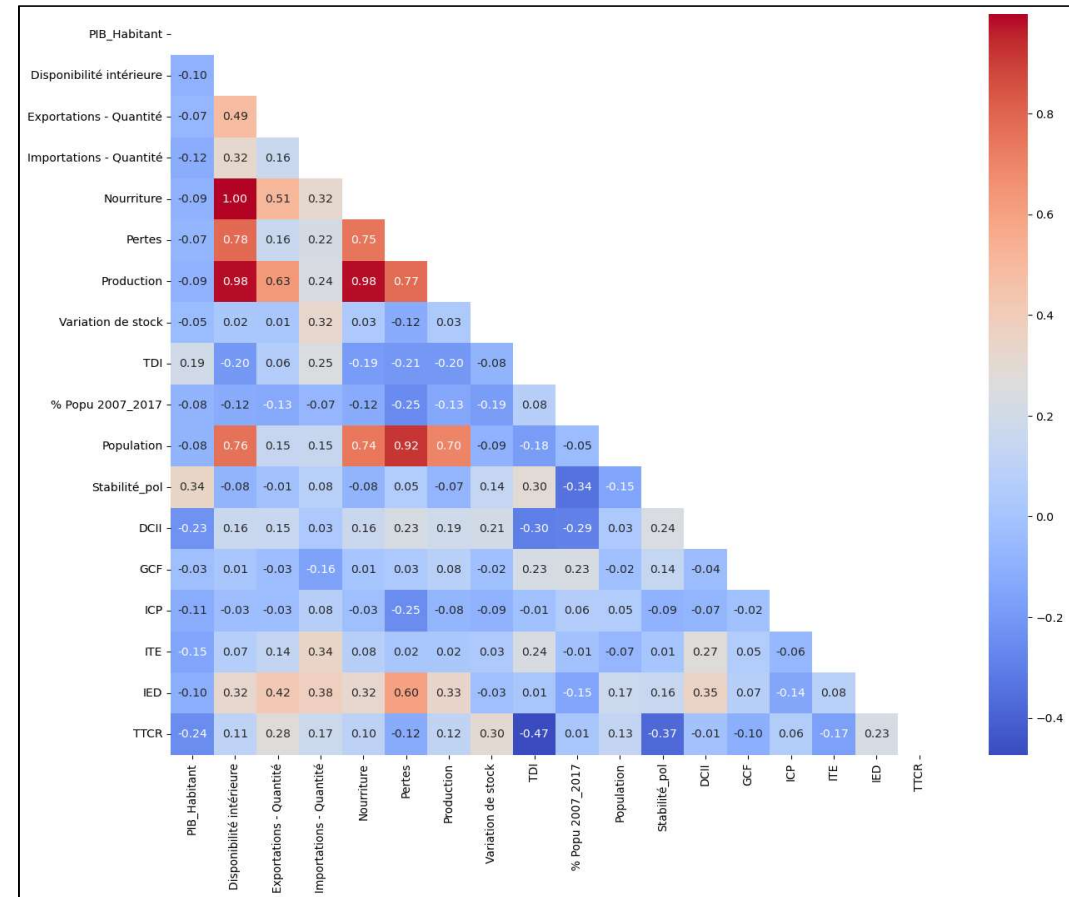
ICP 48 non-null

# Jointures

La jointure des 11 fichiers de données résulte un jeu de données de 18 variables et on a gardé 14 pour nos analyses on supprime les variables qui contiennent beaucoup de valeurs nulles.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 153 entries, 0 to 152
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Pays                                153 non-null    object
1   PIB_Habitant                        152 non-null    float64
2   Disponibilité intérieure            151 non-null    float64
3   Exportations - Quantité             119 non-null    float64
4   Importations - Quantité             151 non-null    float64
5   Nourriture                          151 non-null    float64
6   Pertes                              60 non-null     float64
7   Production                          149 non-null    float64
8   Variation de stock                  150 non-null    float64
9   TDI                                 151 non-null    float64
10  % Popu 2007_2017                   152 non-null    float64
11  Population                          153 non-null    float64
12  Stabilité_pol                       153 non-null    float64
13  Groupe_revenu                      152 non-null    object
14  DCII                               48 non-null     float64
15  GCF                                45 non-null     float64
16  ICP                                48 non-null     float64
17  ITE                                43 non-null     float64
18  IED                                148 non-null    float64
19  TTCR                               48 non-null     float64
dtypes: float64(18), object(2)
memory usage: 25.1+ KB
```

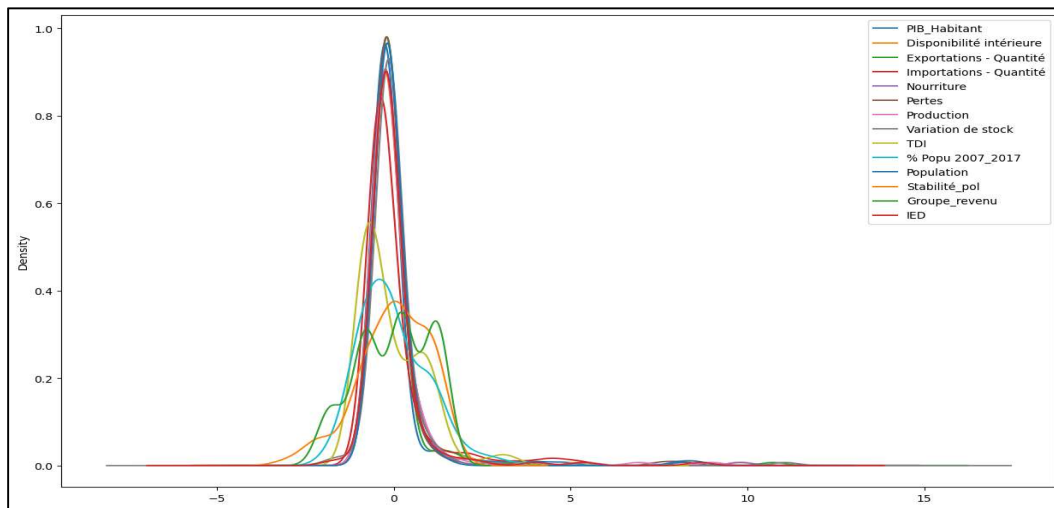
Infos du jeu de données final



Corrélations des variables



Pour pouvoir manipuler nos données on les a nettoyé,  
analyser et préparer  
On est arrivé a un jeu de données de 14 variables dans  
152 pays.  
Et on est trouvé avec le besoin de réduire les 14  
variables



Avant la réduction de la taille de notre  
jeu de données on est passée par la  
standardisation des données puis leur  
centrage

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 152 entries, 0 to 152
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   PIB_Habitant                          152 non-null    float64
1   Disponibilité intérieure              152 non-null    float64
2   Exportations - Quantité              152 non-null    float64
3   Importations - Quantité              152 non-null    float64
4   Nourriture                           152 non-null    float64
5   Pertes                               152 non-null    float64
6   Production                           152 non-null    float64
7   Variation de stock                   152 non-null    float64
8   TDI                                  152 non-null    float64
9   % Popu 2007_2017                    152 non-null    float64
10  Population                           152 non-null    float64
11  Stabilité_pol                         152 non-null    float64
12  Groupe_revenu                        152 non-null    float64
13  IED                                  152 non-null    float64
dtypes: float64(14)
memory usage: 17.8 KB
```

```
array([[ -0.30491628, -0.29338597, -0.21900994, ..., -3.13615696,
        -1.78386012, -0.28833946],
       [ -0.30275988,  0.85667735, -0.06325739, ..., -0.28243666,
         0.21722282, -0.20923547],
       [ -0.2689725 , -0.29896661, -0.22679757, ...,  0.46496628,
         0.21722282, -0.25006047],
       ...,
       [ -0.30313795, -0.29171194, -0.22420169, ...,  0.20450768,
        -0.78331865, -0.24671974],
       [ -0.30272827, -0.28278374, -0.21900994, ..., -0.76938099,
        -0.78331865, -0.27826334],
       [ -0.29068297,  0.11284697, -0.20603056, ...,  1.00853205,
         0.21722282, -0.2434694 ]])
```



```
array([0.33, 0.51, 0.61, 0.69, 0.78, 0.84, 0.89, 0.93, 0.96, 0.98, 1. , 1. , 1. , 1. ])
```

# Réduction du dimension

L'Analyse en Composantes Principales ou ACP, c'est une méthode statistique de réduction de dimension d'un jeu de données qui permet de simplifier et d'analyser des données multidimensionnelles en identifiant les relations et les tendances principales.

```
array([0.31, 0.5 , 0.62,  
0.71, 0.78, 0.85, 0.9 ,  
0.94, 0.97, 0.99, 1. , 1. ])
```

Code python d'application:

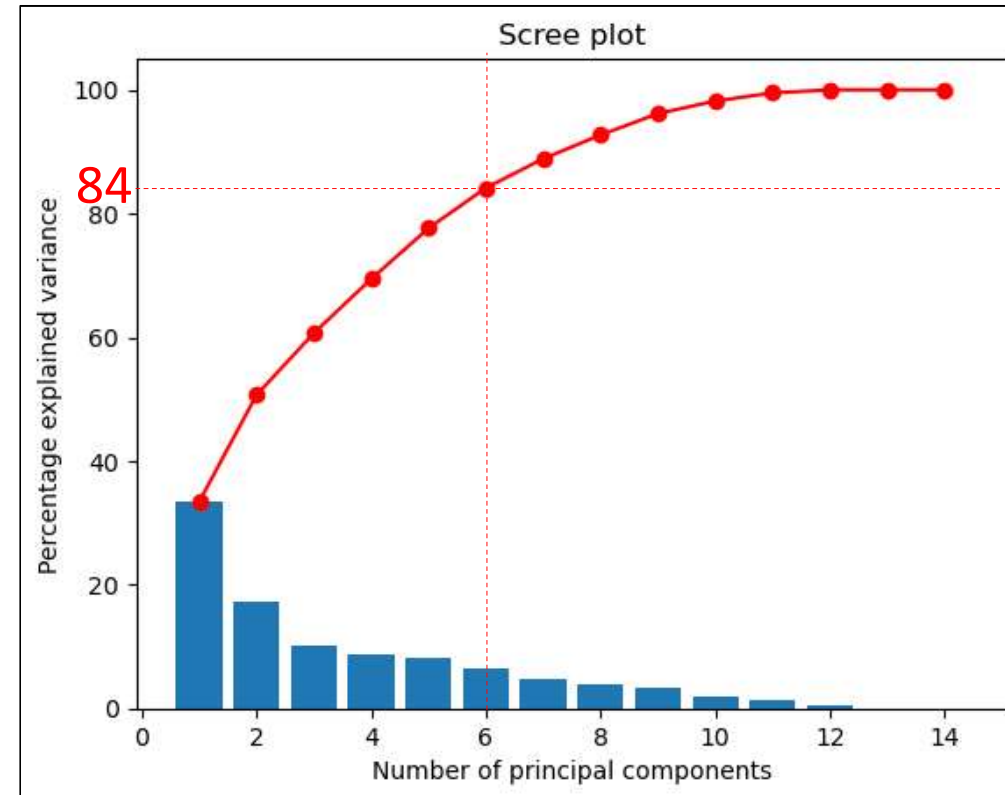
```
# Select the number of principal components  
we will return  
num_components = 14  
  
# Create the PCA model  
pca = PCA(n_components=num_components)  
  
# Fit the model with the standardised data  
pca.fit(X_scaled)
```

on calcule la somme cumulative qui va aider dans le choix du nombre de composantes principales à choisir

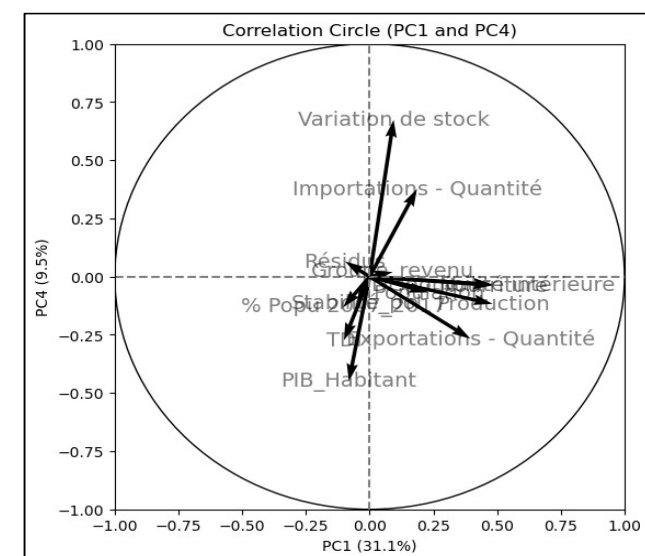
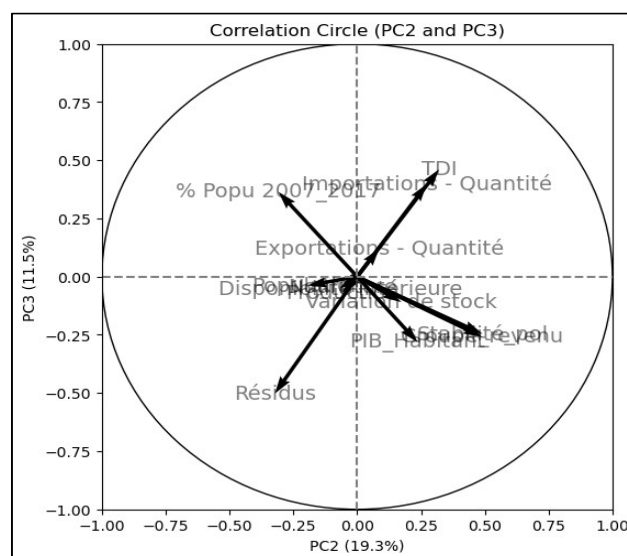
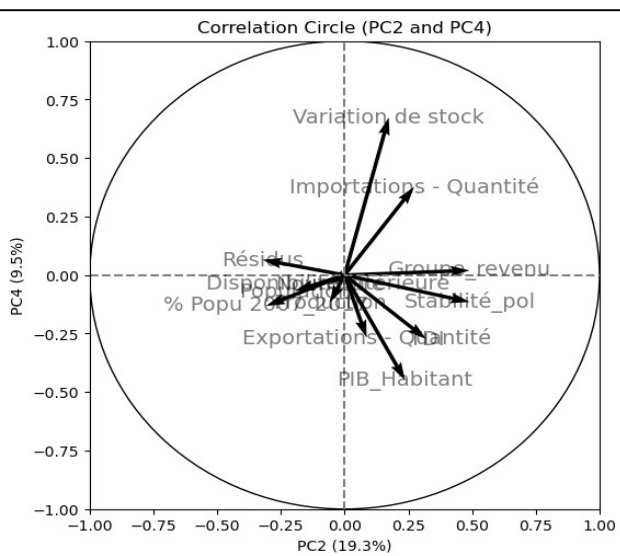
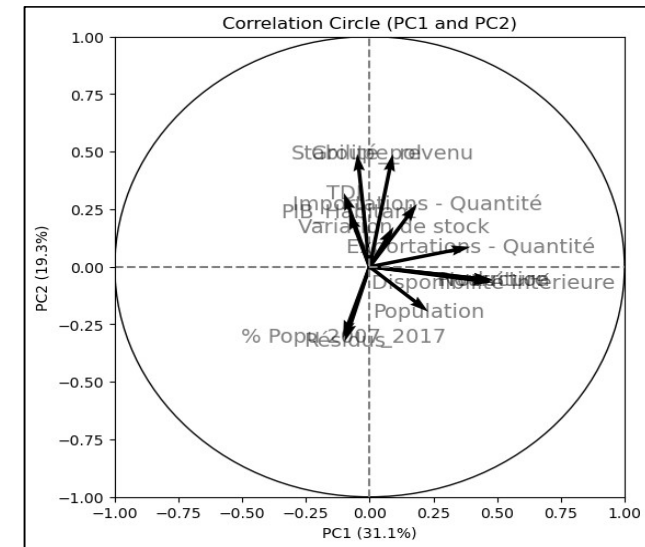
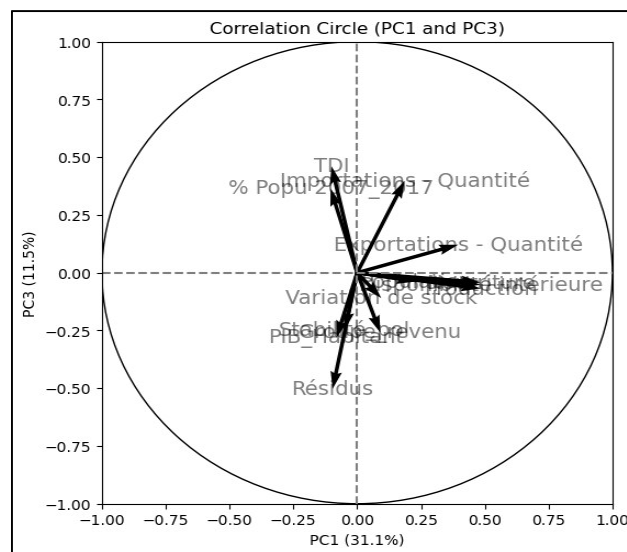
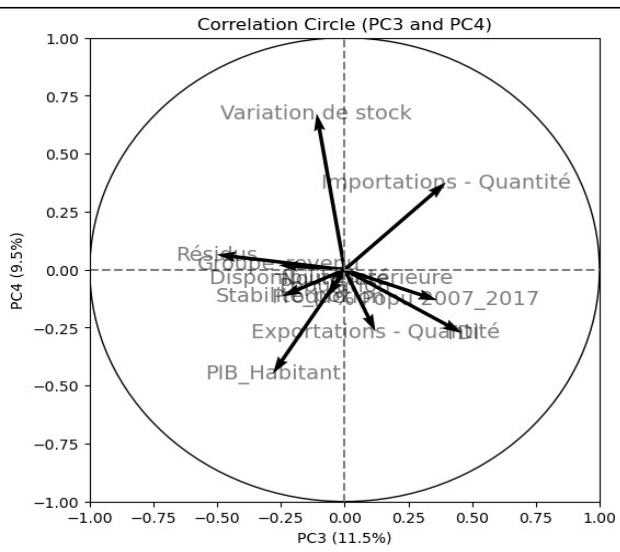
# Identification du nombre de composantes principales

On visualise le graphique de coude pour afficher les valeurs propres des composantes principales identifier le nombre optimal de composantes principales à retenir pour réduire la dimension des données tout en conservant l'essentiel de l'information

Depuis le graphe de coude et selon le critère de Kaiser nous choisissons de garder les 6 premières composantes principales qui représentent environ 84% de l'information originale.



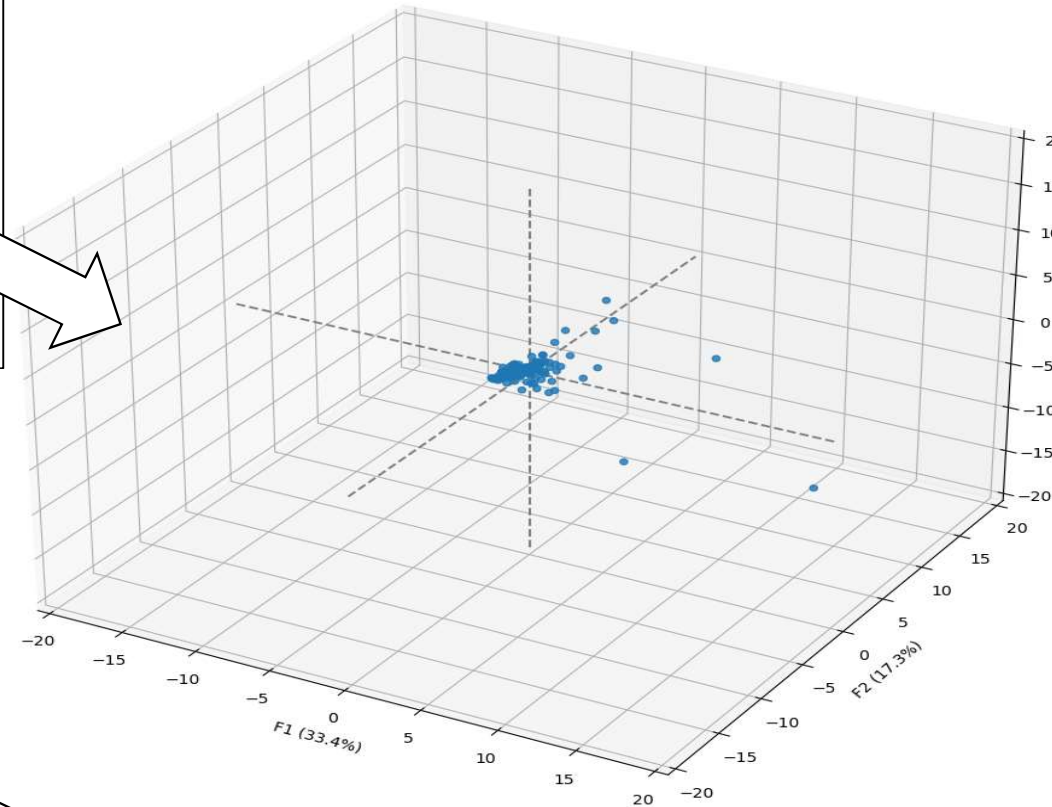
# Visualisation de répartition des variables sur les plan principales



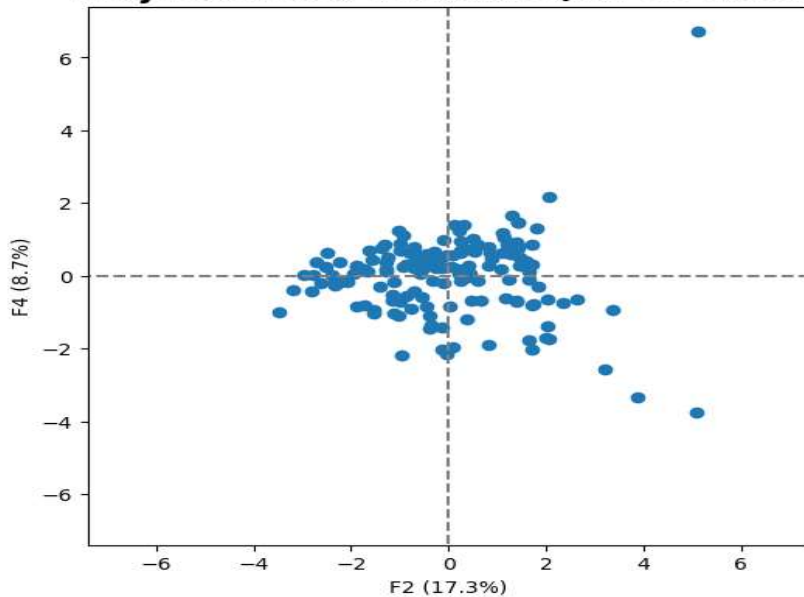
# Visualisations

On a représenté les  
individus réduits dans  
l'espace  
CP2XCP3XCP4

Projection des individus (sur F1 et F2 et F3)



Projection des individus (sur F2 et F4)



Représentation des individus sur le  
plan CP2XCP4

# La décomposition des individus avec des algorithmes de clustering (K-Means)

## 1- K-means:

C'est un algorithme de regroupement de données qui partitionne un ensemble de données en clusters (groupes) en fonction de la similarité des données. Il recherche les centres de clusters en minimisant la somme des distances entre les points de données et leur centre de cluster le plus proche.

### 1- Application K-Means

```
# Creation d'un modèle de clustering k-means
kmeans = KMeans(init='random', n_clusters=10, n_init=10)

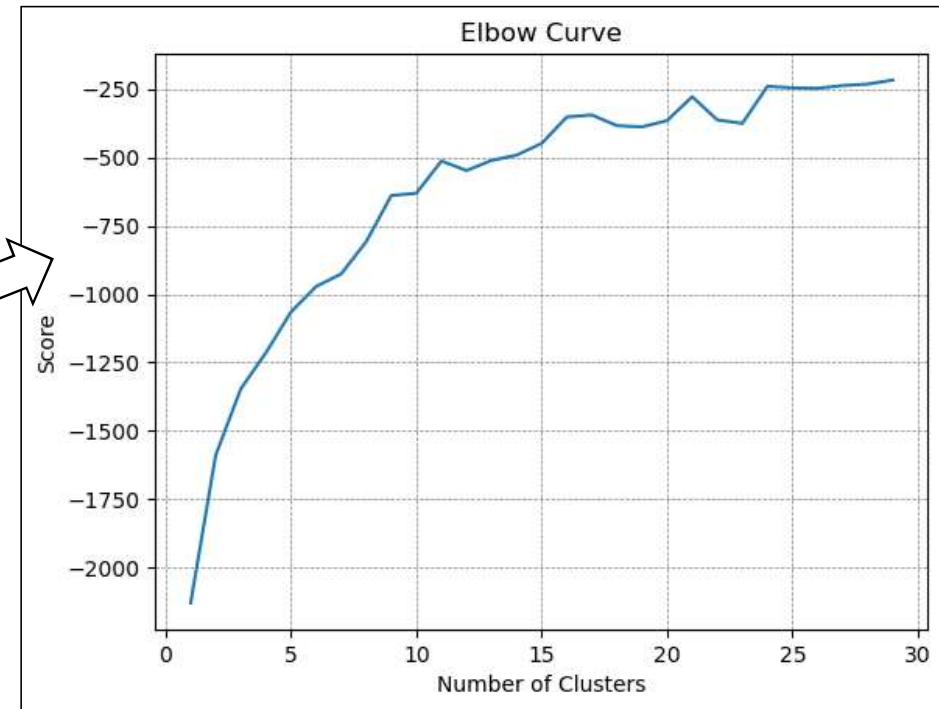
# Fit the data to the model
kmeans.fit(X_scaled)

# Determinion des individus de chaque cluster
clustersKmeans = kmeans.predict(X_scaled)
```

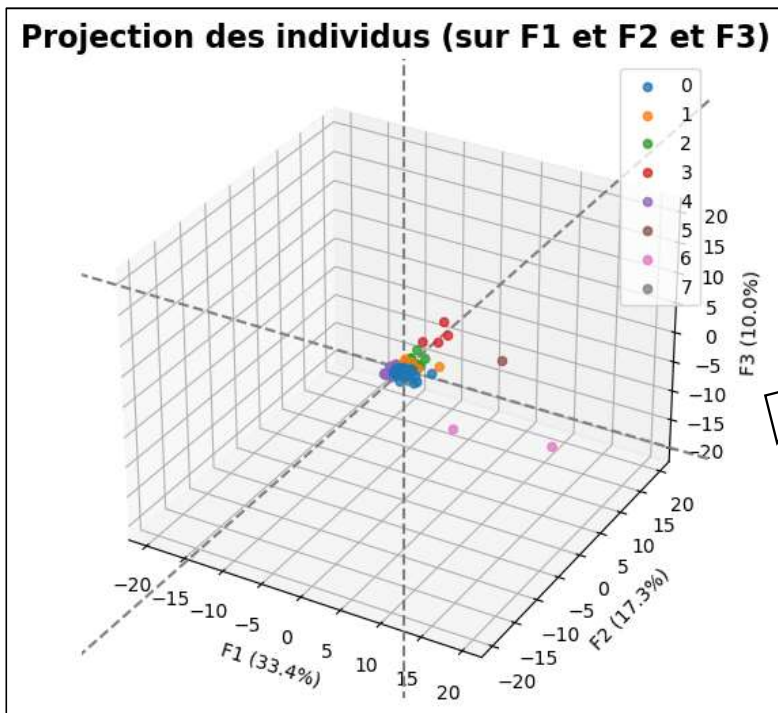
# Visualisations

## 2- choix du nombre des clusters

Selon le graphe de coude on a choisit de partitionner nos données en 6 clusters (groupes)



Graphe de coude

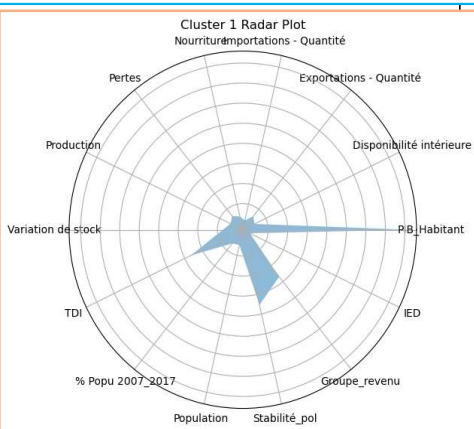
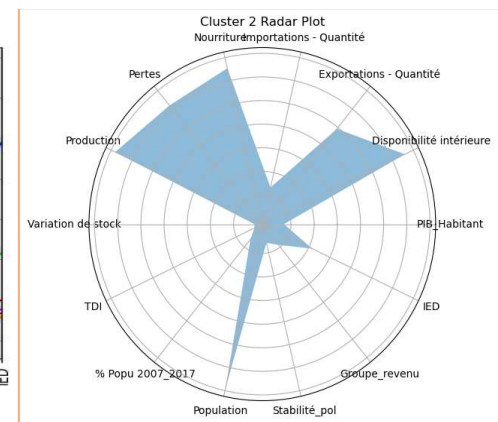
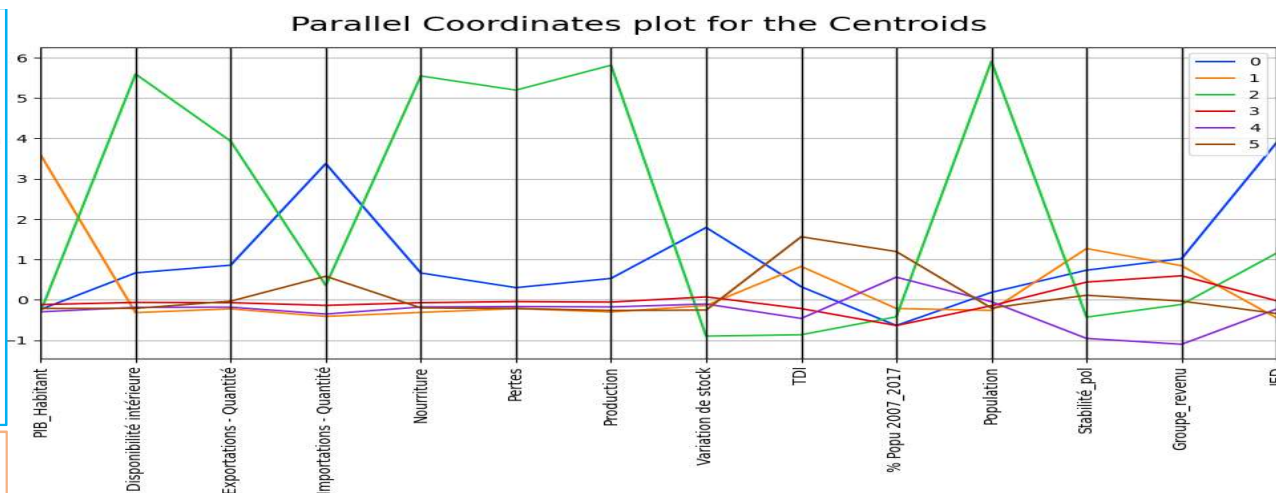
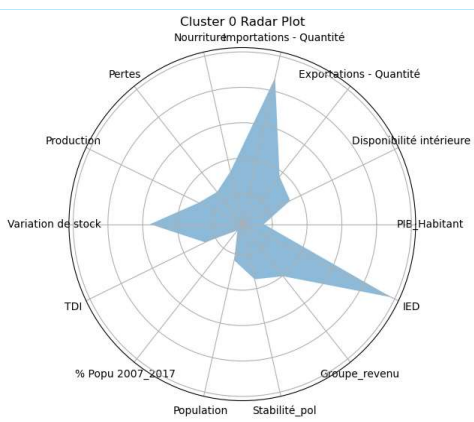


Représentation des 7 cluster sue l'espace 3D (F2XF4XF3)

Nombres d'individus dans chaque groupe selon K-means

cluster	5	: 54
cluster	1	: 68
cluster	0	: 5
cluster	6	: 14
cluster	2	: 8
cluster	4	: 1
cluster	3	: 2





Sur les graphiques en radars:

#### Le cluster 0:

Regroupe les pays avec une grande dépendance à l'importation et peu de disponibilité intérieur.

#### Le cluster 1:

Regroupe les pays qui ont un PIB par habitant important, et qui sont plus ou moins stable politiquement.

#### Le cluster 2:

Regroupe les pays avec une faible dépendance à l'importation avec un revenu et stabilité politique faibles

#### Le cluster 3:

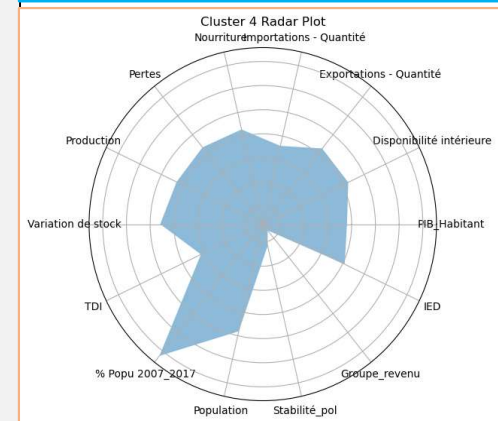
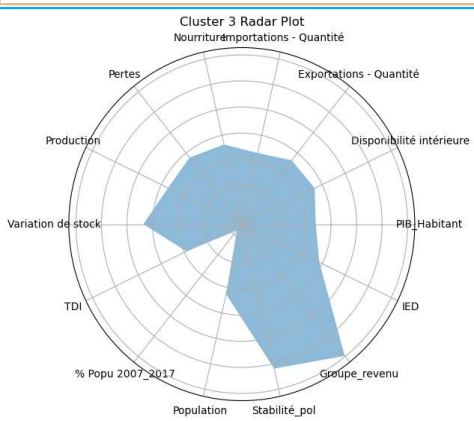
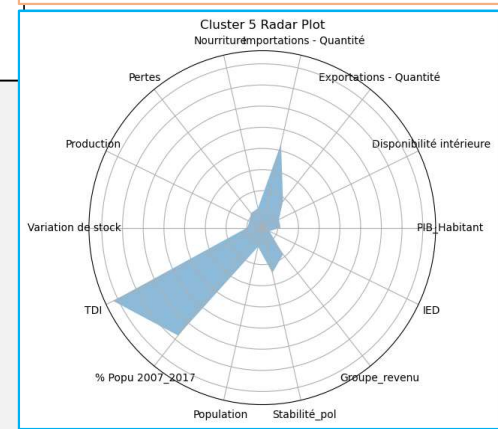
Regroupe les pays avec peu de dépendance à l'importation et une évolution de population faible, stabilité politique importante.

#### Le cluster 4:

Regroupe les pays avec une dépendance à l'importation plus ou moins importante mais faible revenus et stabilité politique faible.

#### Le cluster 5:

Regroupe les pays qui dépendent fortement à l'importation et ont une évolution de population importante.





# La décomposition des individus avec des algorithmes de clustering (ACH)

## 3- ACH:

L'ACH (Agglomerative Hierarchical Clustering ou Algorithme de Clustering Hiérarchique) est une méthode de regroupement de données qui commence par considérer chaque point comme un cluster individuel, puis fusionne progressivement les clusters les plus similaires jusqu'à obtenir un seul cluster contenant tous les points.

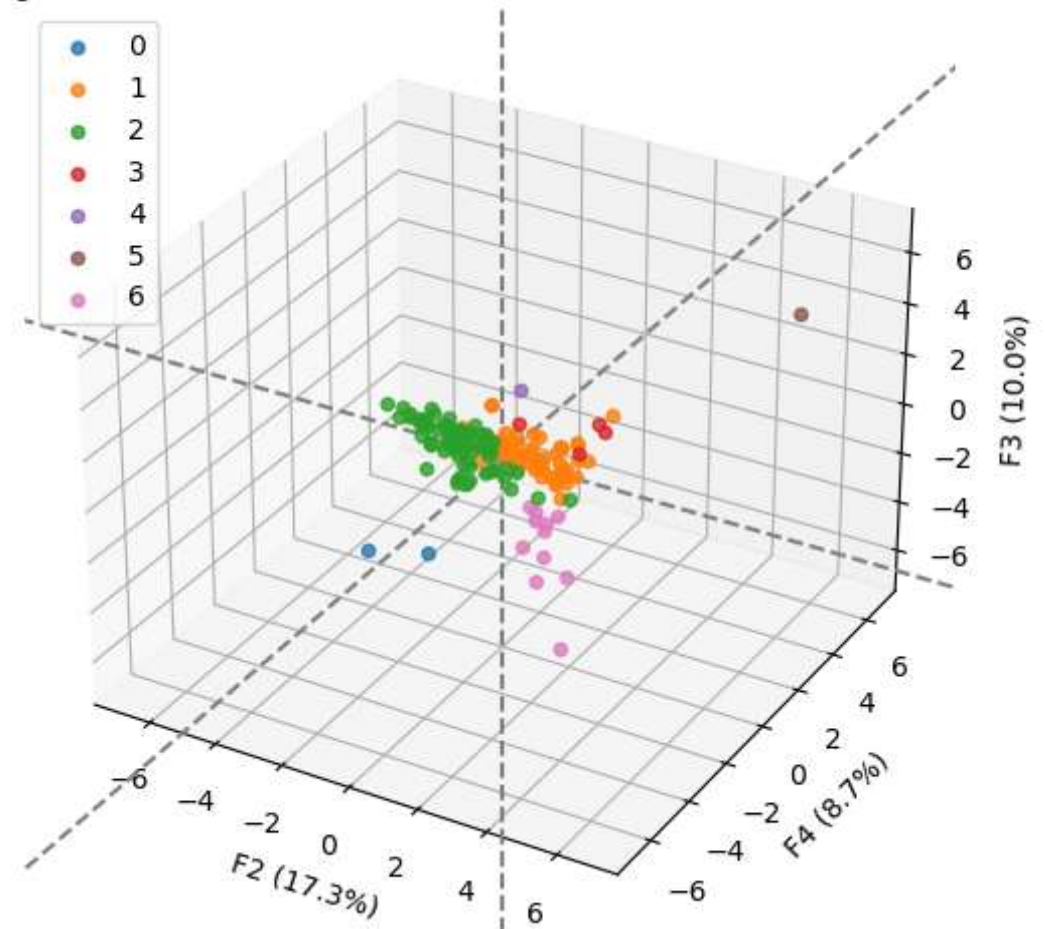
### 1- Application ACH

```
# Fit the data to the model and determine  
which clusters each data point belongs to:  
hiercluster.set_params(n_clusters=7)  
  
clusters = hiercluster.fit_predict(X_scaled)  
np.bincount(clustersACH)
```

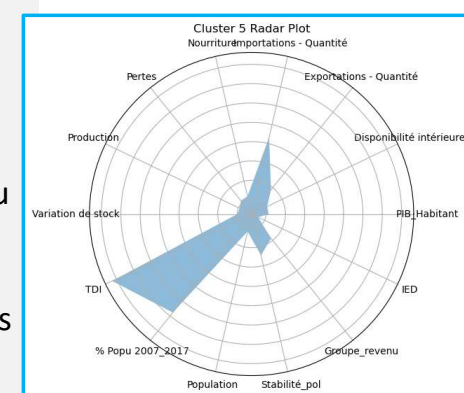
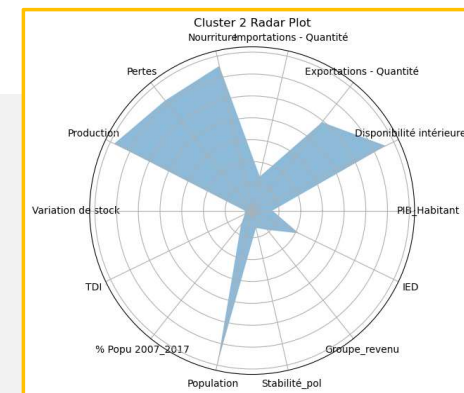
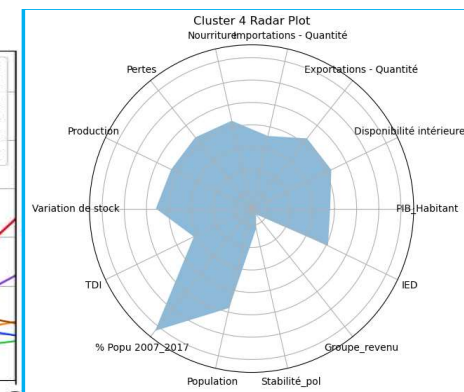
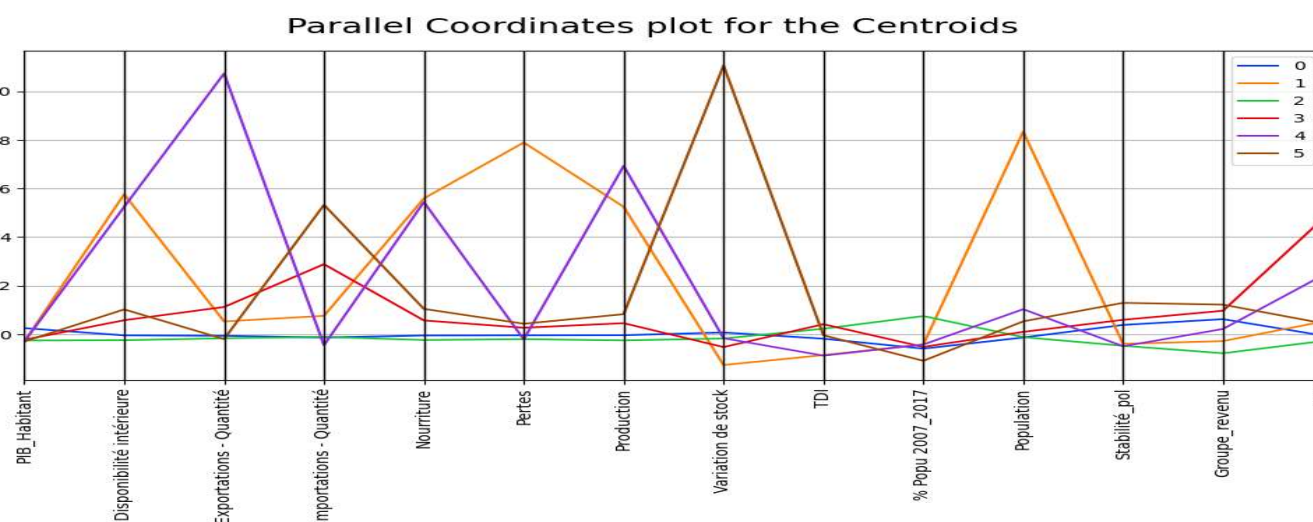
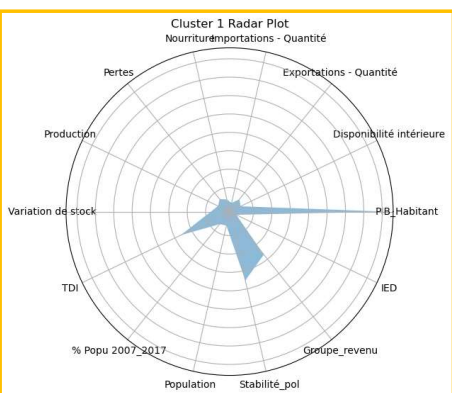
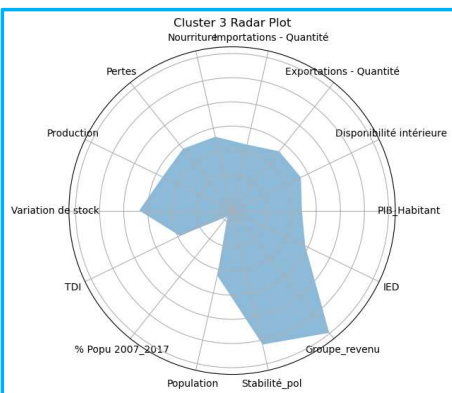
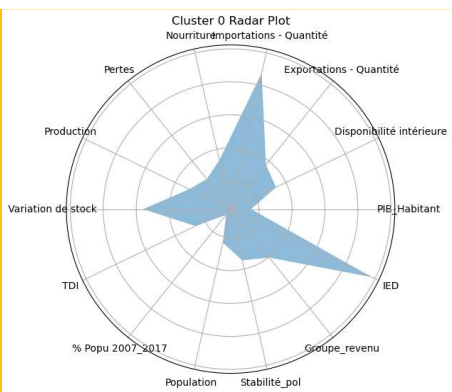
Représentation des 7 cluster sur l'espace  
3D (F2XF4XF3)

cluster	2	:	51
cluster	1	:	16
cluster	3	:	52
cluster	4	:	3
cluster	6	:	5
cluster	0	:	21
cluster	7	:	2
cluster	5	:	2

Projection des individus (sur F2 et F4 et F3)



Nombres d'individus dans chaque  
groupe selon Kmeans



Sur les graphiques de radars on remarque :

#### Le cluster 0:

Regroupe les pays avec un indice de dépendance à l'importation important mais un revenu assez faible.

#### Le cluster 1:

Regroupe les pays qui ont un PIB très important mais faible importation

#### Le cluster 2:

Regroupe les pays production et disponibilité importantes et importation faible.

#### Le cluster 3:

Regroupe les pays qui ont une stabilité politique et revenus importants et dépendent un peu sur l'importation

#### Le cluster 4:

Regroupe les pays avec une importante évolution de population et une dépendance plus ou moins importante.

#### Le cluster 5:

Regroupe les pays avec une dépendance à l'importation et une évolution de population très importantes.

# Recommandations

## Selon les résultats de la K-Means

### Cluster 0:

Allemagne  
Japon  
Mexique  
Pays-Bas  
Suisse

### Cluster 2:

Bahamas  
Chine, RAS de Macao  
Grenade  
Dominique  
Islande  
Luxembourg  
Saint-Kitts-et-Nevis

### Cluster 6:

Angola  
Arabie saoudite  
Belgique  
Chine, RAS de Hong-Kong  
Émirats arabes unis  
Gabon  
Gambie  
Îles Salomon  
Koweït  
Maldives  
Mongolie  
Oman  
Timor-Leste  
Vanuatu

## Selon les résultats de la CAH

### Cluster 6:

Afrique du Sud  
Argentine  
Azerbaïdjan  
Belize  
Cambodge  
Colombie  
Salvador  
Équateur  
Guatemala  
Fédération de Russie  
Eswatini  
Indonésie  
Kazakhstan  
Israël  
Philippines  
Pérou  
Paraguay  
Thaïlande  
Sri Lanka  
République dominicaine  
République démocratique populaire lao

### Cluster 3:

Brésil  
Chine, continentale  
Inde

Azerbaïdjan  
Argentine  
Afrique du Sud  
Ouzbékistan  
Nicaragua  
Népal  
Myanmar  
Mexique  
Maroc  
Malaisie  
Guatemala  
Équateur  
Salvador  
Colombie  
Cambodge  
Belize  
Viet Nam  
Ukraine  
Turquie  
Turkménistan  
Tunisie

## Conclusion

Pour conclure, après analyser les deux méthodes de clustering ressortissent des résultats plutôt similaires avec des groupements de pays qui vont être facilement identifiables et qui seront différents par leurs caractéristiques. A l'intérieur de ces groupements, on trouvera des pays qui présenteront des similitudes entre eux.

Pour l'export de notre viande de volaille on recommande de cibler des pays tirés des clusters résultants par les 2 méthodes et en se basant sur nos expériences et connaissances autour des marchés dans le monde et ce que je recommande et de choisir parmi les pays suivants:

<b>Asie:</b>	<b>Europe:</b>	<b>Amériques:</b>	<b>Afrique:</b>
Arabie saoudite	Belgique	Saint-Kitts-et-Nevis	Angola
Émirats arabes unis	Suisse	Brésil	Libéria
Chine, continentale	Pays-Bas	Dominique	Lesotho
Chine, RAS de Hong-Kong	Allemagne	Grenade	Namibie
Chine, RAS de Macao	Luxembourg	Mexique	
Inde	Islande		
Koweït			
Maldives			
Japon			

### Je recommande :

Belgique, Suisse, Pays-Bas, Allemagne, Luxembourg, Islande, Arabie Saoudite, Emirats Arabes Unis, Chine, RAS Hong Kong, Inde, Japon, Brésil, Mexique, Angola, Namibie, Lesotho