

# Etude de marché



**La poule qui chante**

# Point abordé pendant la présentation :

- Contexte du projet
- Préparation des données
  - Dataframe dispo\_alimentaire
  - Dataframe population
  - Dataframe exploitable pour l'analyse
- Classification hiérarchique
  - Dendrogramme
  - Affichage des groupes de variables
  - Sélection du/des groupes
- ACP
  - Eboulis des valeurs propres
  - Cercle des corrélations
  - Projection sur plan factoriel
- Kmeans
  - Métrique de l'inertie
  - Affichage des groupes de variables
  - Sélection du/des groupes
- Résultats et recommandations

# Contexte du projet

Faire une analyse des pays pour choisir ceux qui sont aptes pour un lancement à l'international.

Il y a 2 fichiers csv à disposition :

- DisponibiliteAlimentaire\_2017.csv : qui contient les informations sur le types de données récupérées, le pays, à quoi est destiné le produit, le produit, l'année, unité et sa valeur et le symbole et sa description.
- Population\_2000\_2018.csv : contient les informations sur les types de données, le pays , la population prit en compte, l'année (Ces données sont aussi identifiées par un code), l'unité et sa valeur, le symbole et sa description et une note.

# Dataframe dispo\_alimentaire

Code Domaine	Domaine	Code zone	Zone	Code Élément	Élément	Code Produit	Produit	Code année	Année	Unité	Valeur	Symbole	Description du Symbole
FBS	Nouveaux Bilans Alimentaire	2	Afghanistan	5511	Production	2511	Blé et produits	2017	2017	Milliers de tonnes	4281.00	S	Données standardisées
FBS	Nouveaux Bilans Alimentaire	2	Afghanistan	5611	Importations - Quantité	2511	Blé et produits	2017	2017	Milliers de tonnes	2302.00	S	Données standardisées
FBS	Nouveaux Bilans Alimentaire	2	Afghanistan	5072	Variation de stock	2511	Blé et produits	2017	2017	Milliers de tonnes	-119.00	S	Données standardisées
FBS	Nouveaux Bilans Alimentaire	2	Afghanistan	5911	Exportations - Quantité	2511	Blé et produits	2017	2017	Milliers de tonnes	0.00	S	Données standardisées
FBS	Nouveaux Bilans Alimentaire	2	Afghanistan	5301	Disponibilité intérieure	2511	Blé et produits	2017	2017	Milliers de tonnes	6701.00	S	Données standardisées
...	...	...	...	...	...	...	...	...	...	...	...	...	...
FBS	Nouveaux Bilans Alimentaire	181	Zimbabwe	5142	Nourriture	2899	Miscellanees	2017	2017	Milliers de tonnes	19.00	S	Données standardisées
FBS	Nouveaux Bilans Alimentaire	181	Zimbabwe	645	Disponibilité alimentaire en quantité (kg/pers...	2899	Miscellanees	2017	2017	kg	1.33	Fc	Donnée calculée
FBS	Nouveaux Bilans Alimentaire	181	Zimbabwe	664	Disponibilité alimentaire (Kcal/personne /jour)	2899	Miscellanees	2017	2017	Kcal/personne/jour	1.00	Fc	Donnée calculée
FBS	Nouveaux Bilans Alimentaire	181	Zimbabwe	674	Disponibilité de protéines en quantité (g/pers...	2899	Miscellanees	2017	2017	g/personne/jour	0.04	Fc	Donnée calculée
FBS	Nouveaux Bilans Alimentaire	181	Zimbabwe	684	Disponibilité de matière grasse en quantité (g...	2899	Miscellanees	2017	2017	g/personne/jour	0.02	Fc	Donnée calculée

Les colonnes gardées sont zone, élément et valeur.

Une sélection dans la colonne produit sur la viande de volailles.

Pivoter la colonne élément car les données sont utiles.

Les données gardées sont la disponibilité alimentaire, disponibilité de protéines en quantité, exportations, importations, et production.

Renommage de la colonne zone en pays.

# Dataframe population

Code Domaine	Domaine	Code zone	Zone	Code Élément	Élément	Code Produit	Produit	Code année	Année	Unité	Valeur	Symbole	Description du Symbole	Note
OA	Séries temporelles annuelles	2	Afghanistan	511	Population totale	3010	Population- Estimations	2000	2000	1000 personnes	20779.953	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	2	Afghanistan	511	Population totale	3010	Population- Estimations	2001	2001	1000 personnes	21606.988	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	2	Afghanistan	511	Population totale	3010	Population- Estimations	2002	2002	1000 personnes	22600.770	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	2	Afghanistan	511	Population totale	3010	Population- Estimations	2003	2003	1000 personnes	23680.871	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	2	Afghanistan	511	Population totale	3010	Population- Estimations	2004	2004	1000 personnes	24726.684	X	Sources internationales sûres	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
OA	Séries temporelles annuelles	181	Zimbabwe	511	Population totale	3010	Population- Estimations	2014	2014	1000 personnes	13586.707	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	181	Zimbabwe	511	Population totale	3010	Population- Estimations	2015	2015	1000 personnes	13814.629	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	181	Zimbabwe	511	Population totale	3010	Population- Estimations	2016	2016	1000 personnes	14030.331	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	181	Zimbabwe	511	Population totale	3010	Population- Estimations	2017	2017	1000 personnes	14236.595	X	Sources internationales sûres	NaN
OA	Séries temporelles annuelles	181	Zimbabwe	511	Population totale	3010	Population- Estimations	2018	2018	1000 personnes	14438.802	X	Sources internationales sûres	NaN

Les colonnes retenues sont zone et valeur.

Sélection sur l'année 2017 car les données sur le dataframe dispo\_alimentaire concerne l'année 2017.

Renomme les colonnes zone en pays et valeur en population pour 1000.

# Dataframe exploitation pour l'analyse

Jointure est faite sur la colonne pays pour lier les 2 dataframes.

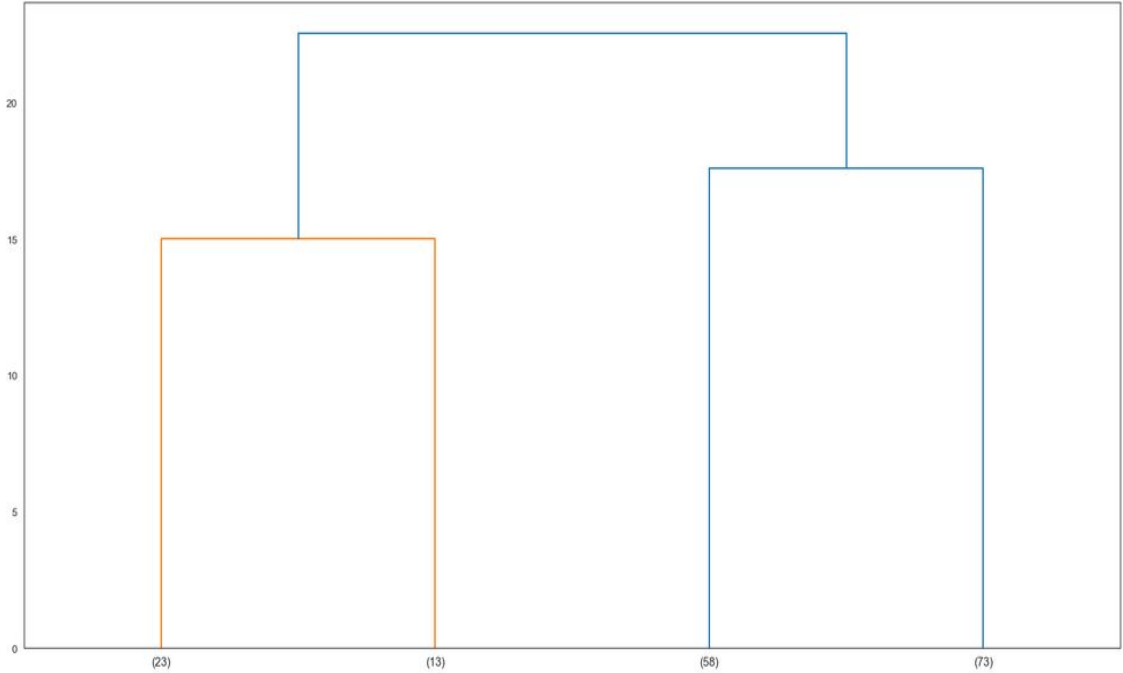
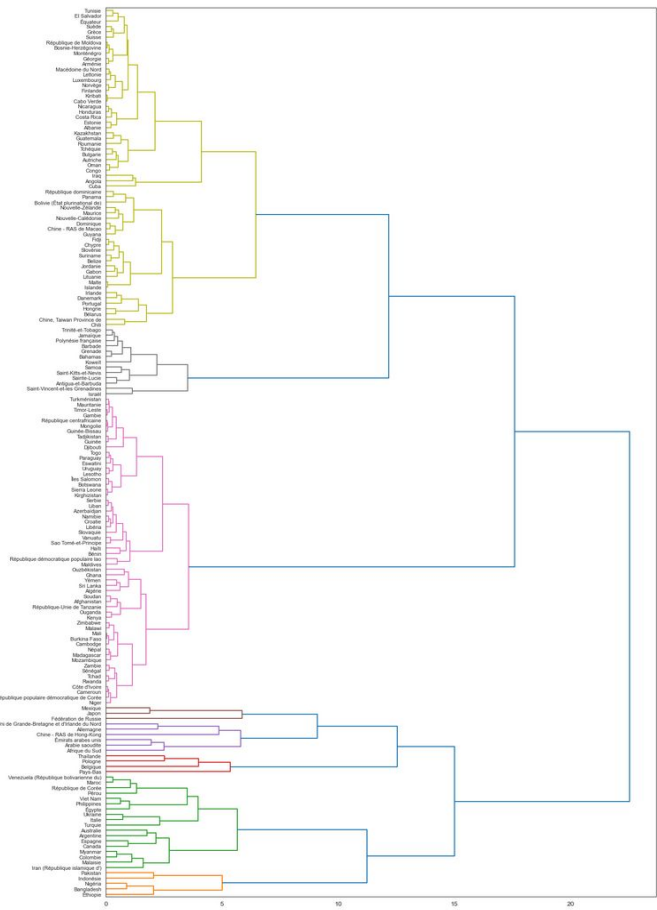
Dans les pays, la France est enlevée.

De plus, la Chine, les Etats-Unis, l'Inde et le Brésil vont être supprimés car pays à forte production de poulet et de population. Il n'est pas pertinent de les viser pour une implantation et leurs données ne sont pas pertinentes pour l'analyse.

La moyenne est utilisé pour combler les valeurs manquantes pour les colonnes exportations, importations et production.

Pays	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	Exportations - Quantité	Importations - Quantité	Production	Population pour 1000
Afghanistan	5.0	0.54	68.07	29.0	28.0	36296.113
Afrique du Sud	143.0	14.11	63.00	514.0	1667.0	57009.756
Albanie	85.0	6.26	0.00	38.0	13.0	2884.169
Algérie	22.0	1.97	0.00	2.0	275.0	41389.189
Allemagne	71.0	7.96	646.00	842.0	1514.0	82658.409
...	...	...	...	...	...	...
Égypte	50.0	4.51	1.00	110.0	1118.0	96442.591
Émirats arabes unis	147.0	14.80	94.00	433.0	48.0	9487.203
Équateur	83.0	6.15	0.00	0.0	340.0	16785.361
Éthiopie	0.0	0.04	68.07	1.0	14.0	106399.924
Îles Salomon	18.0	1.51	0.00	6.0	0.0	636.039

# Dendrogramme en 4 groupes



# Analyse des groupes

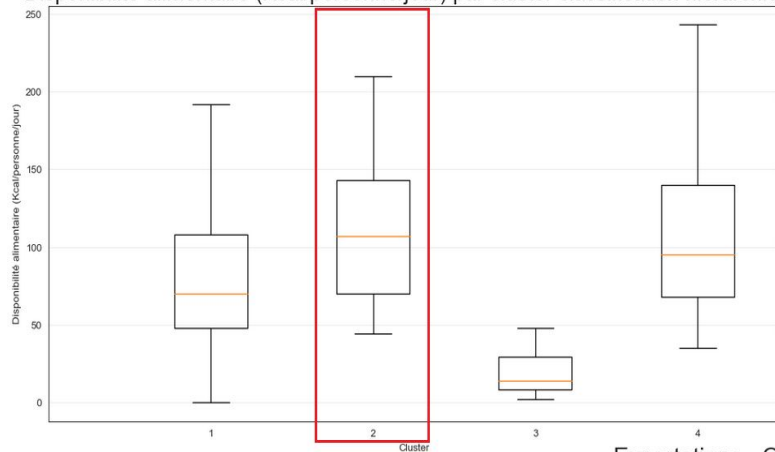
Le choix des pays intéressant pour l'export de poulet sont les pays qui présente les critères suivants :

- Une disponibilité alimentaire et de protéine haute, cela montre une grande consommation de poulet.
- Une exportation basse car les pays qui exportent le plus de poulet, sont les pays qui n'en ont pas besoin.
- Une importation et une production haute car cela montre les pays qui ont le besoin de poulet.
- Une population importante.

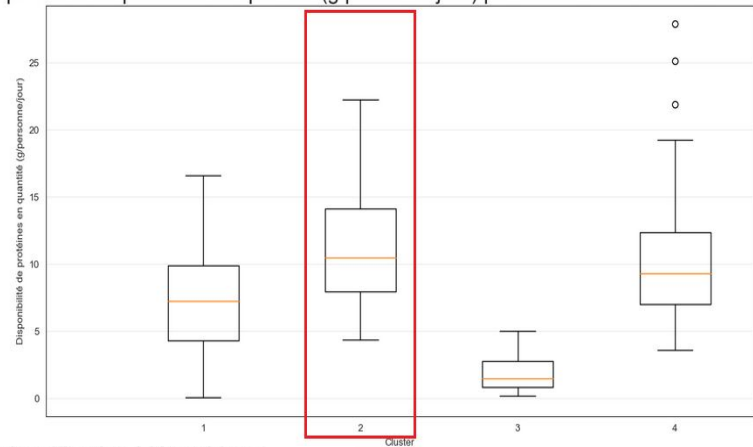


# Affichage des groupes de variables

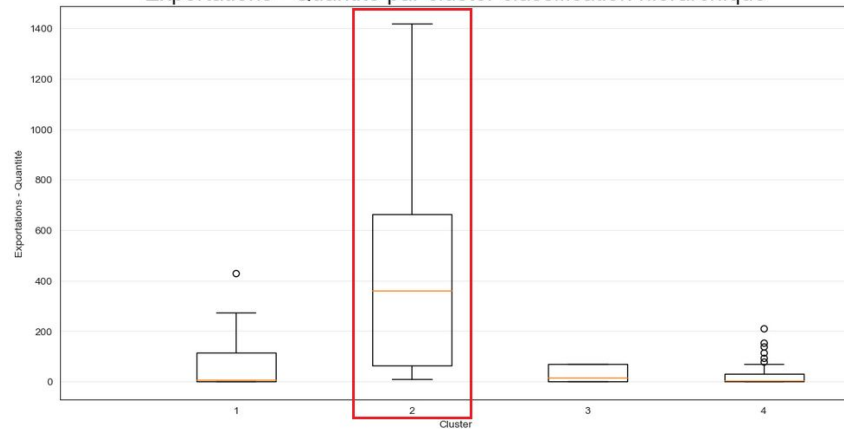
Disponibilité alimentaire (Kcal/personne/jour) par cluster classification hiérarchique



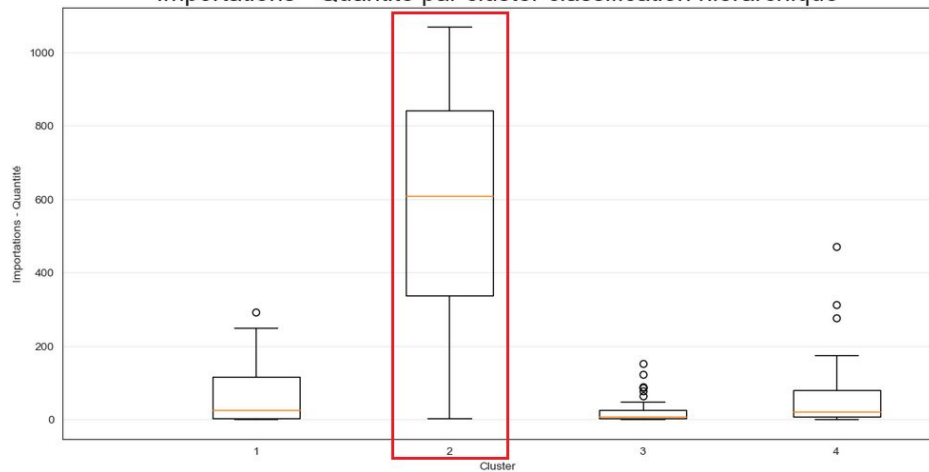
Disponibilité de protéines en quantité (g/personne/jour) par cluster classification hiérarchique



Exportations - Quantité par cluster classification hiérarchique



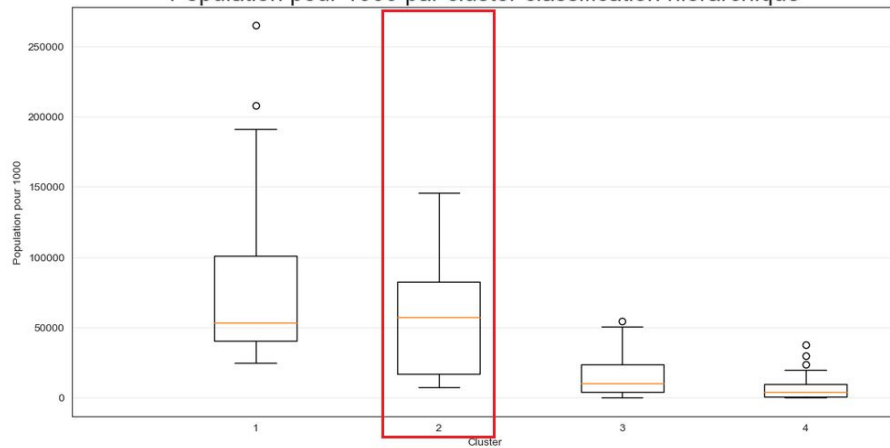
Importations - Quantité par cluster classification hiérarchique



Production par cluster classification hiérarchique



Population pour 1000 par cluster classification hiérarchique



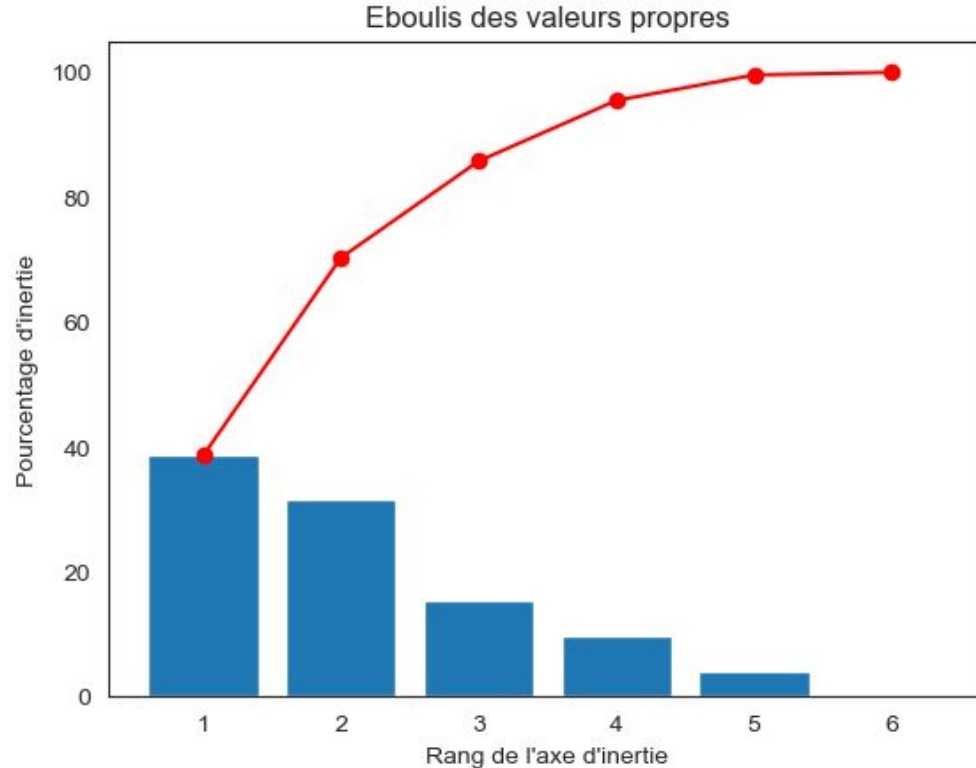
# Sélection du groupe par classification hiérarchique

Le groupe 2 est celui qui correspond le plus aux critères.

Voici les pays contenus dans ce groupe :

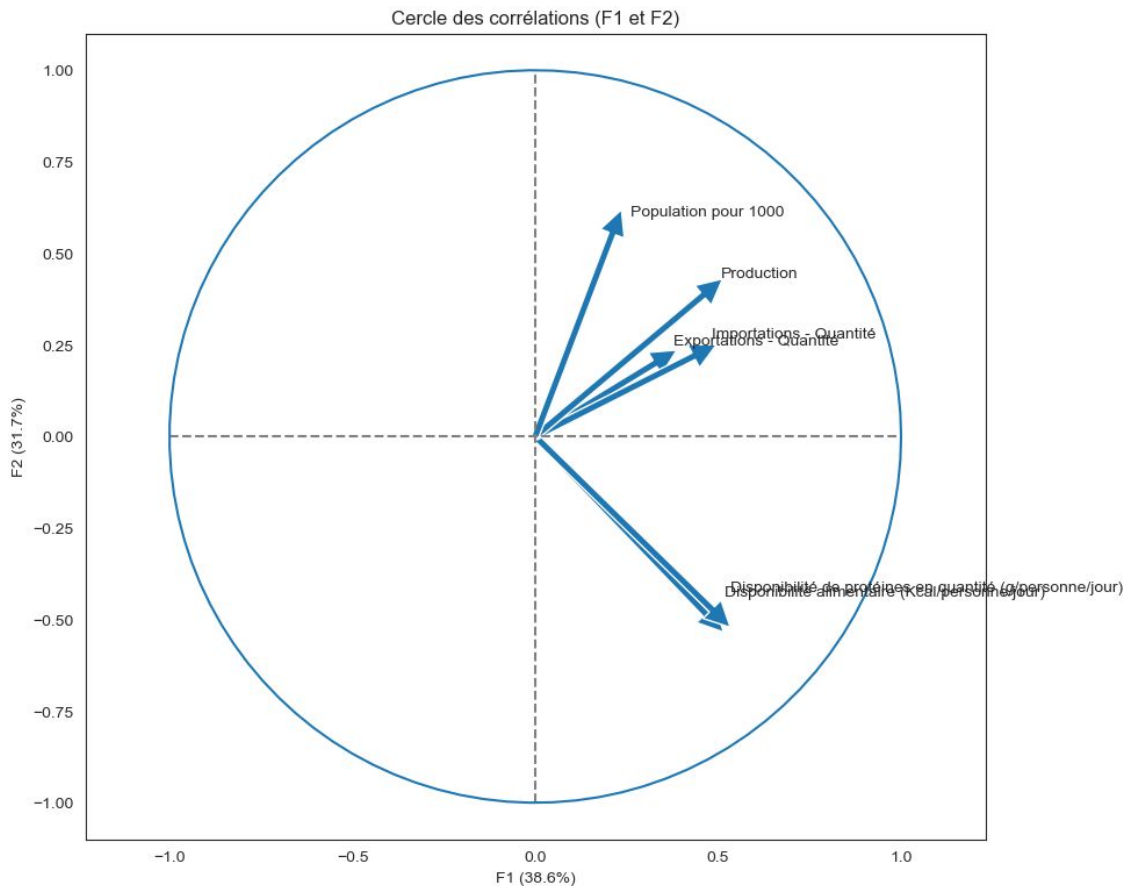
```
'Afrique du Sud', 'Allemagne', 'Arabie saoudite', 'Belgique',  
'Chine - RAS de Hong-Kong', 'Fédération de Russie', 'Japon',  
'Mexique', 'Pays-Bas', 'Pologne',  
"Royaume-Uni de Grande-Bretagne et d'Irlande du Nord", 'Thaïlande',  
'Émirats arabes unis'], dtype=object)
```

# ACP - Eboulis des valeurs propres



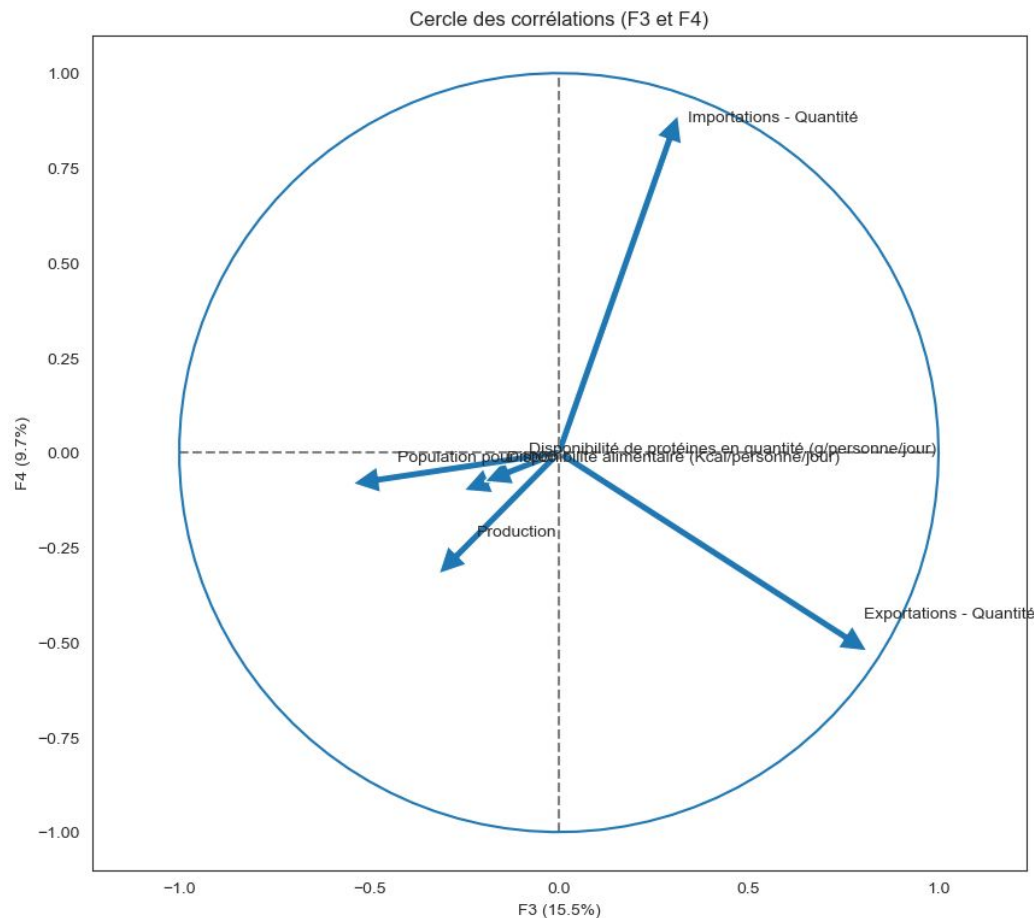
Les 3 premières composantes captent 85,78 % (38,64 + 31,67 + 15,47) de la variance.

# ACP - Cercles des corrélations



En F1, toutes les variables sont corrélées, de plus, la disponibilité alimentaire et la disponibilité de protéines en quantité sont fortement corrélées entre elles.

Pour la composante F2, la population, production, importation et exportation sont décorrélées par rapport à la disponibilité alimentaire et en protéines.

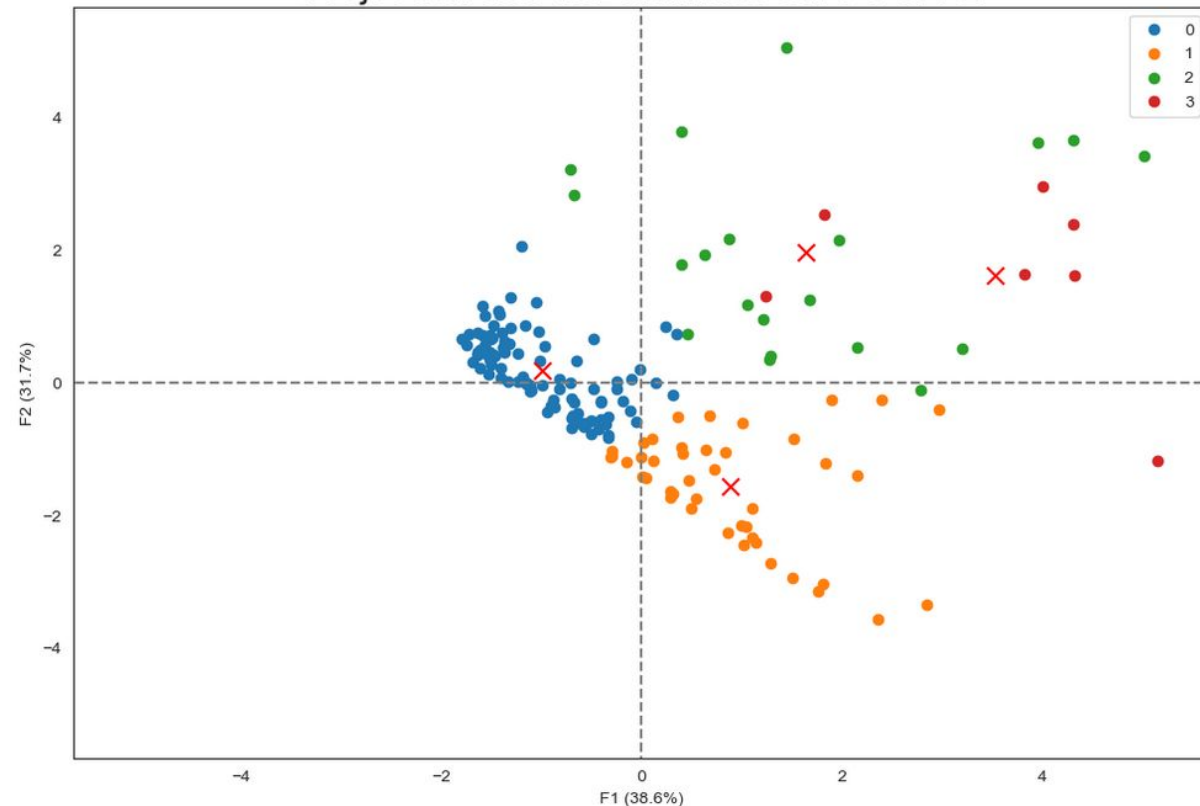


En F3, l'importation et l'exportation sont bien représentées contrairement aux restes des variables.

Pour la composante F4 l'importation et l'exportation sont décorélées entre elles.

# ACP - Projection sur plan factoriel

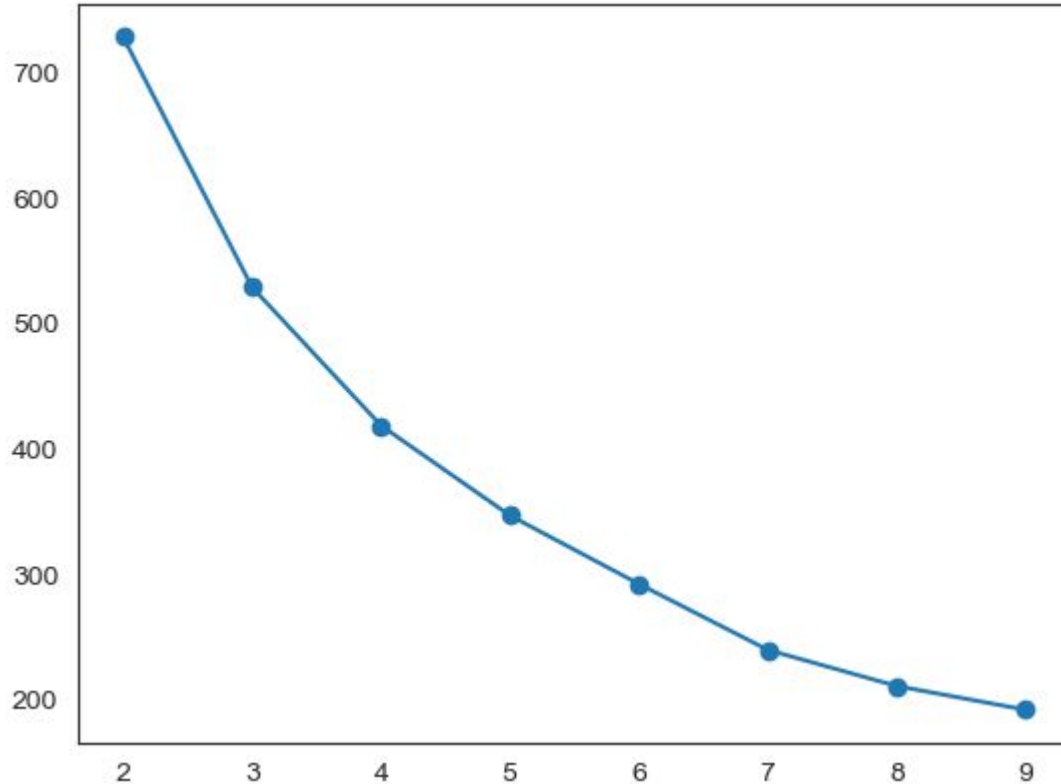
Projection des 167 individus sur F1 et F2



La projection nous permet de mieux représenter les groupes pour permettre une sélection.

Les croix rouges montrent les centroïdes des 4 clusters Kmeans.

# Métrique de l'inertie



On remarque que l'inertie diminue rapidement entre les clusters 3 et 4.

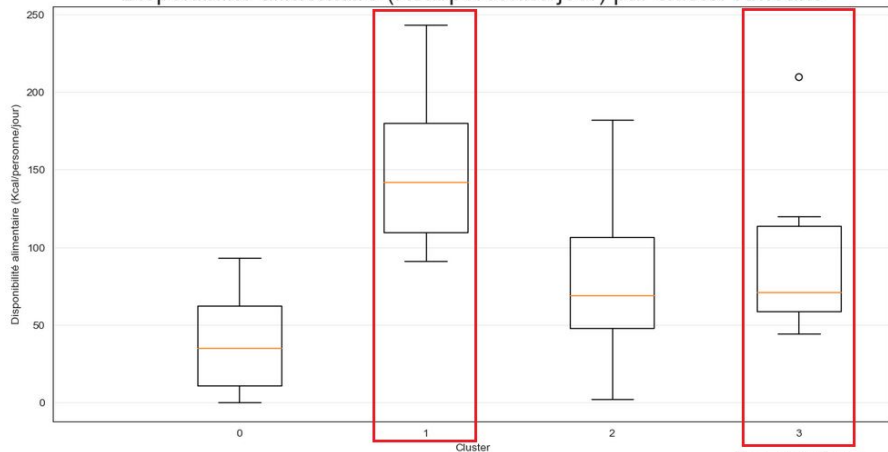
De plus, un nombre de clusters à 5 peuvent être aussi un choix.

Le choix va se porter sur un nombre de clusters à 4 car c'est là où la cassure est la plus importante.

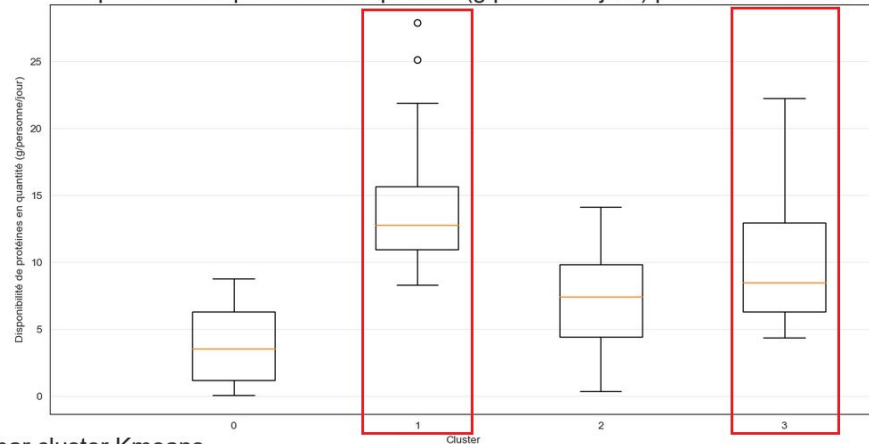


# Affichage des groupes de variables Kmeans

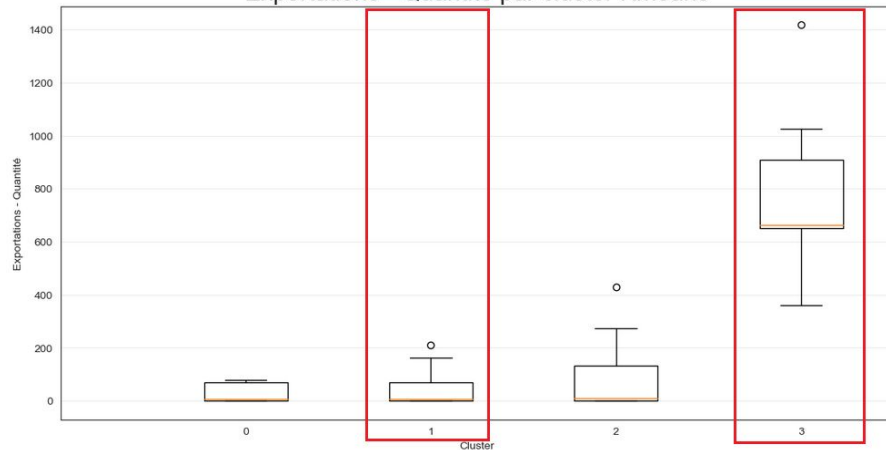
Disponibilité alimentaire (Kcal/personne/jour) par cluster Kmeans



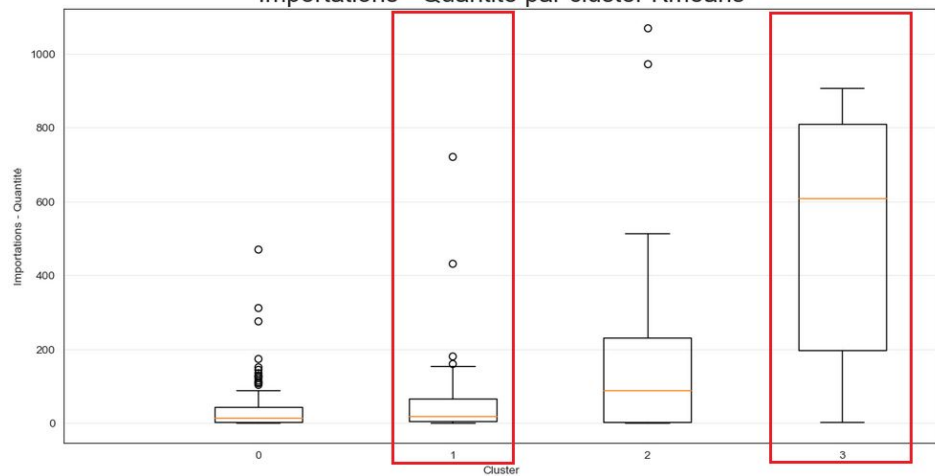
Disponibilité de protéines en quantité (g/personne/jour) par cluster Kmeans



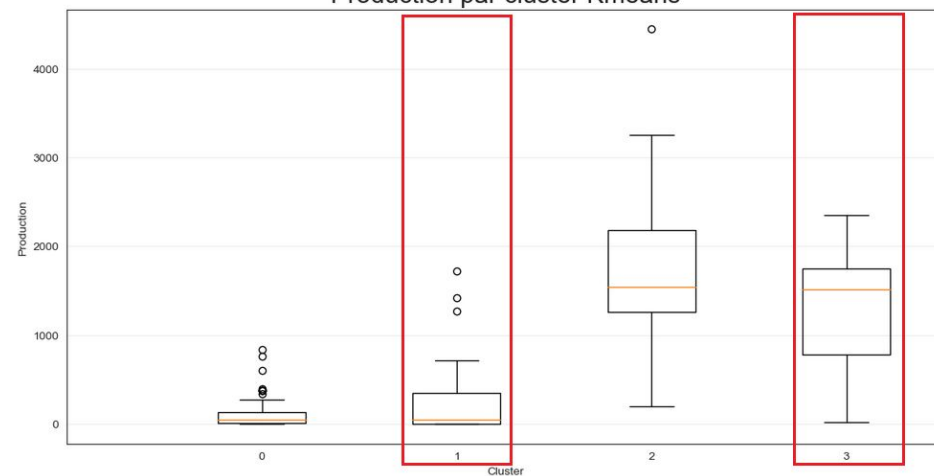
Exportations - Quantité par cluster Kmeans



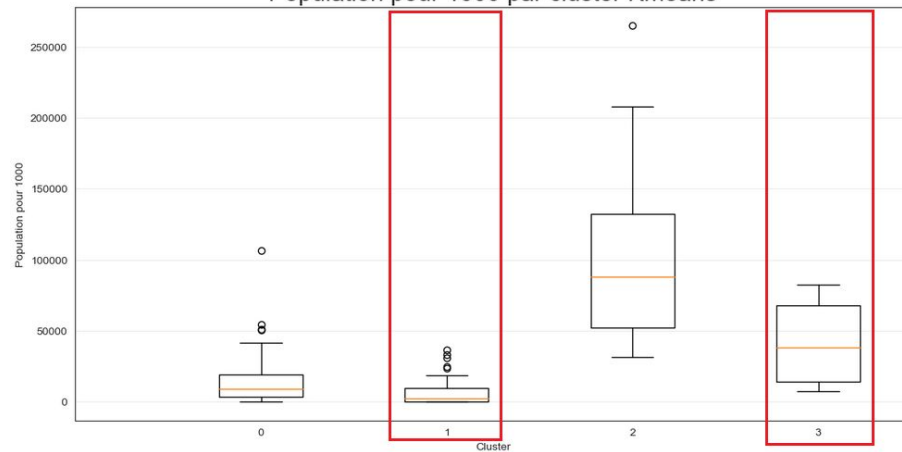
Importations - Quantité par cluster Kmeans



Production par cluster Kmeans



Population pour 1000 par cluster Kmeans



# Sélection des groupes par Kmeans

En prenant en compte les mêmes critères de sélection que pour les groupes par classification hiérarchique, les groupes 1 et 3 sont ceux qui correspondent le plus à ces critères.

Voici les pays contenus dans ces groupes :

```
'Allemagne', 'Antigua-et-Barbuda', 'Arabie saoudite', 'Australie',  
'Bahamas', 'Barbade', 'Belgique', 'Belize',  
'Bolivie (État plurinational de)', 'Biélorus', 'Canada', 'Chili',  
'Chine - RAS de Hong-Kong', 'Chine - RAS de Macao',  
'Chine, Taiwan Province de', 'Chypre', 'Danemark', 'Dominique',  
'Fidji', 'Gabon', 'Grenade', 'Guyana', 'Hongrie', 'Irlande',  
'Islande', 'Israël', 'Jamaïque', 'Jordanie', 'Koweït', 'Lituanie',  
'Malaisie', 'Malte', 'Maurice', 'Nouvelle-Calédonie',  
'Nouvelle-Zélande', 'Panama', 'Pays-Bas', 'Pologne',  
'Polynésie française', 'Portugal',  
'Royaume-Uni de Grande-Bretagne et d'Irlande du Nord',  
'République dominicaine', 'Saint-Kitts-et-Nevis',  
'Saint-Vincent-et-les Grenadines', 'Sainte-Lucie', 'Samoa',  
'Suriname', 'Thaïlande', 'Trinité-et-Tobago',  
'Émirats arabes unis'], dtype=object)
```

# Résultats et recommandations

Pour choisir les pays, on met en comparaison les dataframes des pays sélectionnés par les 2 méthodes. Les pays qui se retrouvent dans les 2 dataframes, vont être les pays choisis.

Voici le liste final des 9 pays :

- Allemagne
- Arabie saoudite
- Belgique
- Chine - RAS de Hong-Kong
- Pays-Bas
- Pologne
- Royaume-Uni de Grande-Bretagne et d'Irlande du Nord
- Thaïlande
- Émirats arabes unis