




Etude de marché sur l'exportation à l'international



Parcours Data Analyste
MARCHAND Emmanuel – 10/2024



Agenda

Partie 1 : Sourcing et préparation des données

Partie 2 : Analyse ACP et Clustering.



Partie 1 :

Sourcing et préparation des données

Collecte des données



Source	Donnée	URL de la source
FAOSTAT	Disponibilité alimentaire	Fourni par Patrick
FAOSTAT	Population	Fourni par Patrick
WTO STATS (OMC)	Importations bilatérales	https://stats.wto.org/ sélectionner Bilateral imports
WTO STATS (OMC)	Taux de douane	https://stats.wto.org/ sélectionner Tariff indicators – applied / MFN agricultural products
World Bank Group	Indice de pénétration des exportations	https://datacatalog.worldbank.org/search/dataset/0064713/Index-Of-Export-Market-Penetration-Export
ILOSTAT	Taux de chômage	https://ilostat.ilo.org/fr/data/# Sélectionner Taux de chôma/plus de données
World Bank group	Niveau de revenu par pays	https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups Sélectionner 'historical classification by income in xls format'

Composantes PESTEL couvertes.

Socioculturel :
préférences alimentaires et consommation.

Politique :
Réglementations commerciales, politiques fiscales et relations internationales.

Economique :
Situation économique, taux de chômage

Préparation et nettoyage des données 1/2



Population :

- Filtre pour conserver **2017**
- Clean des colonnes inutiles à l'analyse
- Le pays représente quel pourcentage de la pop totale mondiale?
- Seulement **12 pays** représentent **60%** de la population mondiale
- Sélection des **100 pays** qui représentent **96,5%** de la population mondiale

Dispo Alimentaire :

- Filtre pour conserver les produits de type '**Viande de volaille**'
- Clean des colonnes inutiles à l'analyse
- On conserve les éléments représentatifs de la composante consommation :
 - Production
 - Importations
 - Dispo intérieure
 - Exportations
 - Pertes
- Chaque élément étant une ligne pour chaque pays, on transforme le dataset (pivot) pour obtenir une seule ligne pays et autant de colonnes qu'il y a d'éléments.



Importations bilatérales :

- Filtre pour conserver le secteur '**Animal products**'
- Filtre pour conserver les importations bilatérales **entre la France et les pays de l'OMC** (code partenaire 250)
- Clean des colonnes inutiles à l'analyse

Taux de douane :

- Filtre pour conserver **2017**
- Filtre pour conserver **le taux de douane simple** (non pondéré, pas de min pas de max)
- Clean des colonnes inutiles à l'analyse

Préparation et nettoyage des données 2/2



Taux de chômage :

- Filtre pour conserver **2017**
- On retient les données **non générées**.
- On fait la moyenne pour **toutes les tranches d'âge**.
- Clean des colonnes inutiles à l'analyse



Niveau de revenu par pays :

- Filtre depuis la source Excel pour conserver uniquement 2017

Indice de de pénétration des exportations :

- Clean des colonnes inutiles à l'analyse.
- Indices fournit par **code ISO3** pays et non pas par libellé de pays.



Food and Agriculture
Organization of the
United Nations



WORLD TRADE
ORGANIZATION



ETAPE 1 : Uniformisation des libellés de zone

- Traduction Anglais vers français
- Transformation des code ISO3 en libellés. **Web scraping** du site Wikipédia.
- Suppression d'agrégation de pays, d'acronyme.
- Alignement des libellés Français sur ceux du fichier dispo alimentaire

ETAPE 2 : Création du fichier d'analyse

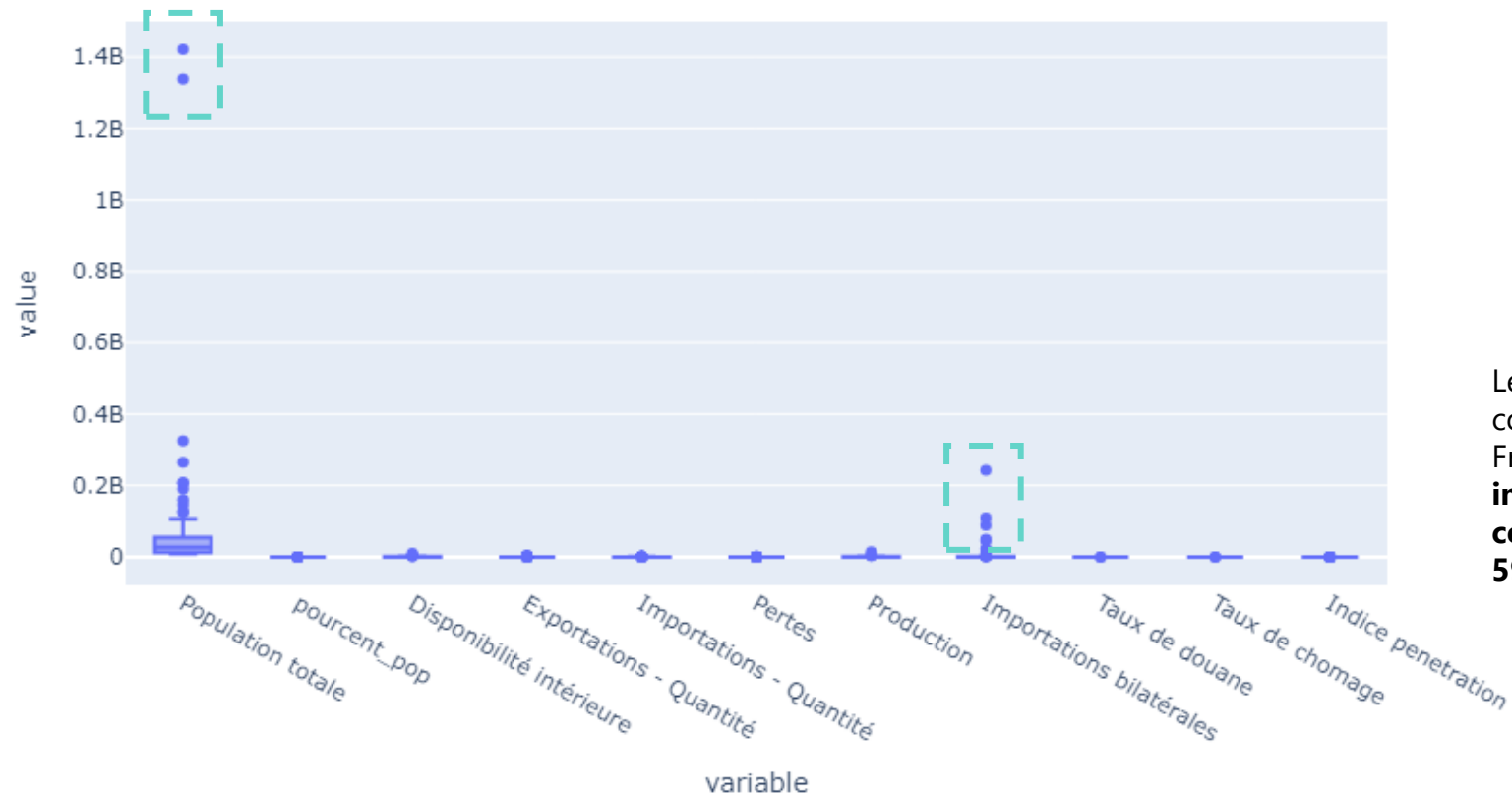
- Merge sur le libellé de pays de tous les fichiers de données, l'un après l'autre.
- Remplacement des NaN **par des zéros**.

Description du jeu de données final

Libellé colonne	Description
Pays	Libellé du pays en français
Population totale	Population totale du pays (nombre exact sans unité)
Pourcent pop	Pourcentage de la population mondiale que représente la population du pays, exprimé en pourcentage.
Disponibilité intérieure	Disponibilité de viande de volaille en tonne.
Exportation quantité	Exportations de viande de volaille en tonne.
Importation quantité	Importation de viande de volaille en tonne.
Pertes	Pertes de viande de volaille en tonne.
Production	Production de viande de volaille en tonne.
Importations bilatérales	Montant en USD des importations bilatérales par catégorie de produit MTN entre chaque pays de l'OMC et la France. Nous avons retenu la catégorie produits animaux. On prend l'hypothèse que plus le montant est élevé plus il y a d'échanges commerciaux avec le pays, car les accords bilatéraux (MTN : Négociations Commerciales Multilatérales) lèvent les obstacles au commerce.
Taux de douane	Taux de douane simple appliqué sur les produits agricoles, exprimé en pourcentage
Taux de chômage	Taux de chômage moyen du pays quel que soit le sexe et pour toutes les tranches d'âges, exprimé en pourcentage
Indice de pénétration	Cet indicateur mesure dans quelle mesure les exportations d'un pays atteignent des marchés déjà établis. Un indice faible indique la présence d'obstacles au commerce qui empêchent les entreprises du pays à exporter (nombre sans unité)
Niveau de revenu du pays	Variable qualitative qui indique le niveau de revenu de chaque pays (classification OMC) : L : Low (revenu faible) ; LM : Lower middle (revenu moyen inférieur) ; UM : Upper middle (Revenu moyen supérieur) ; H : High (Revenu élevé)

Exploration du jeu de données final

Boîtes à moustache des variables du jeu de données



Population totale,
deux outliers connus :
Chine et Inde.

Les échanges commerciaux entre la France et ses **10 plus importants partenaires commerciaux** représente **599 419 872 USD**.

Exploration du jeu de données final

```
RangeIndex: 100 entries, 0 to 99
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Zone                                  100 non-null    object
1   Population totale                     100 non-null    float64
2   pourcent_pop                          100 non-null    float64
3   Disponibilité intérieure              100 non-null    float64
4   Exportations - Quantité              100 non-null    float64
5   Importations - Quantité              100 non-null    float64
6   Pertes                               100 non-null    float64
7   Production                           100 non-null    float64
8   Importations bilatérales              100 non-null    float64
9   Taux de douane                       100 non-null    float64
10  Taux de chômage                       100 non-null    float64
11  Indice penetration                    100 non-null    float64
12  niveau_revenu                         100 non-null    object
dtypes: float64(11), object(2)
```

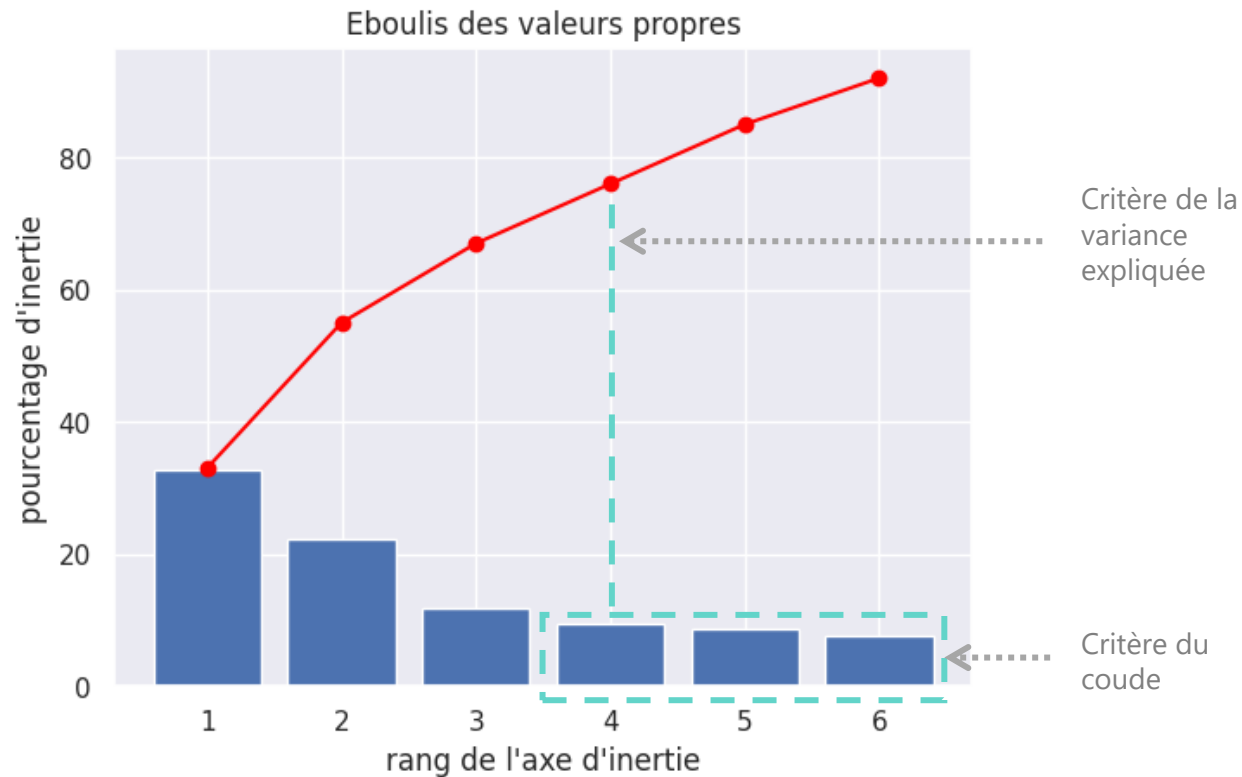
En résumé :

- Toutes les données sont de **l'année 2017**
- Pour chaque variable nous avons une valeur **zéro** si **l'information** n'est pas présente pour le pays.
- Il y a **100 pays** et **12 variables** d'analyse
- Quelques outliers qui s'expliquent
- Le jeu de données couvre les **trois premiers** axes de **l'analyse PESTEL**



Partie 2 : ACP et Clustering

Analyse en composantes principales (ACP)



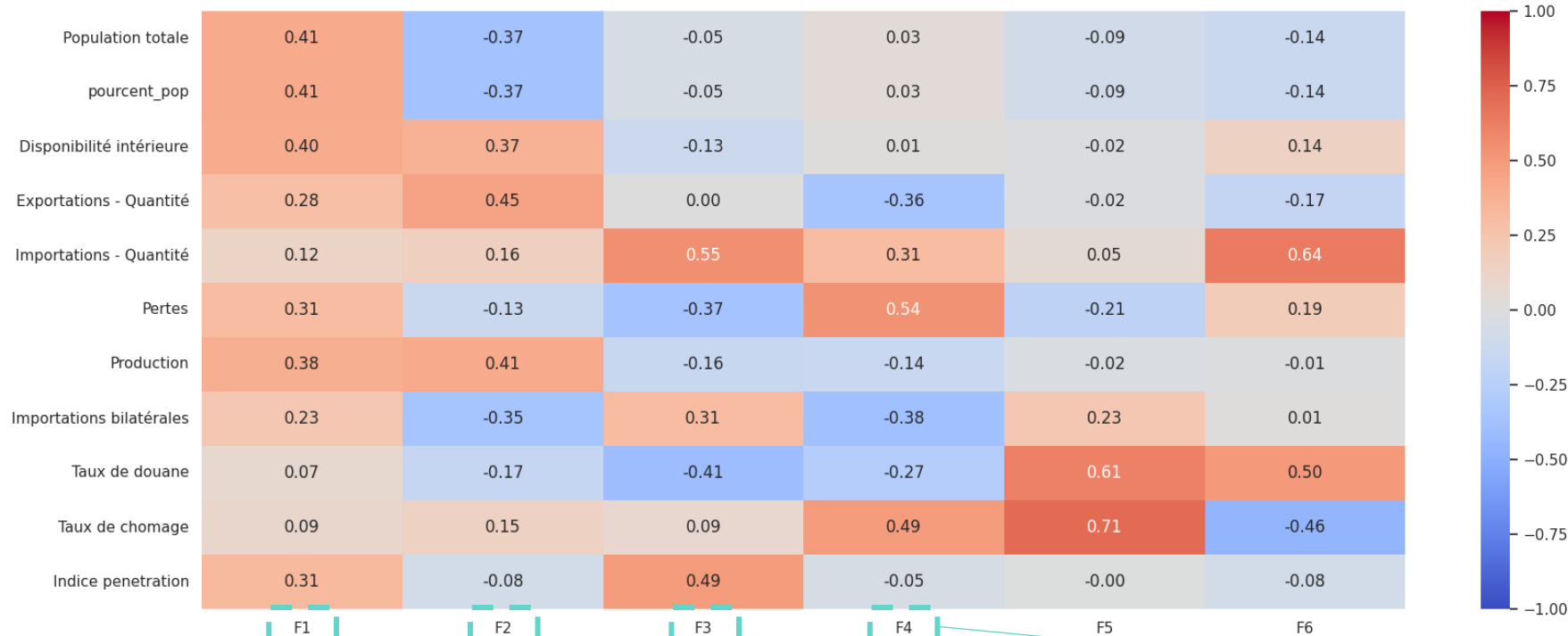
On a en bleu le pourcentage d'inertie (la variance de chaque nouvelle composante) , et en rouge la variance cumulée.

A partir de nos **12 variables**, l'**ACP** nous propose **6 variables synthétiques**.



Nous retenons les **variables synthétiques 1 à 4**

Corrélations entre les variables



HeatMap des corrélations :

Les variables sont-elles corrélées ?

- 1 très forte corrélation
- -1 anti-corrélation

Variable synthétique F1 :

Composition possible : pop totale, pourcent pop, dispo intérieure, importations , pertes

Hypothèse : Besoin du pays en viande de volaille.

Variable synthétique F2 :

Composition possible : Disponibilité intérieure, Exportations, Production

Hypothèse : Capacité du pays à satisfaire sa demande en viande de volaille.

Variable synthétique F3 :

Composition possible : Importations, Importations bilatérales, Indice pénétration, taux de douane (anti-corrélé)

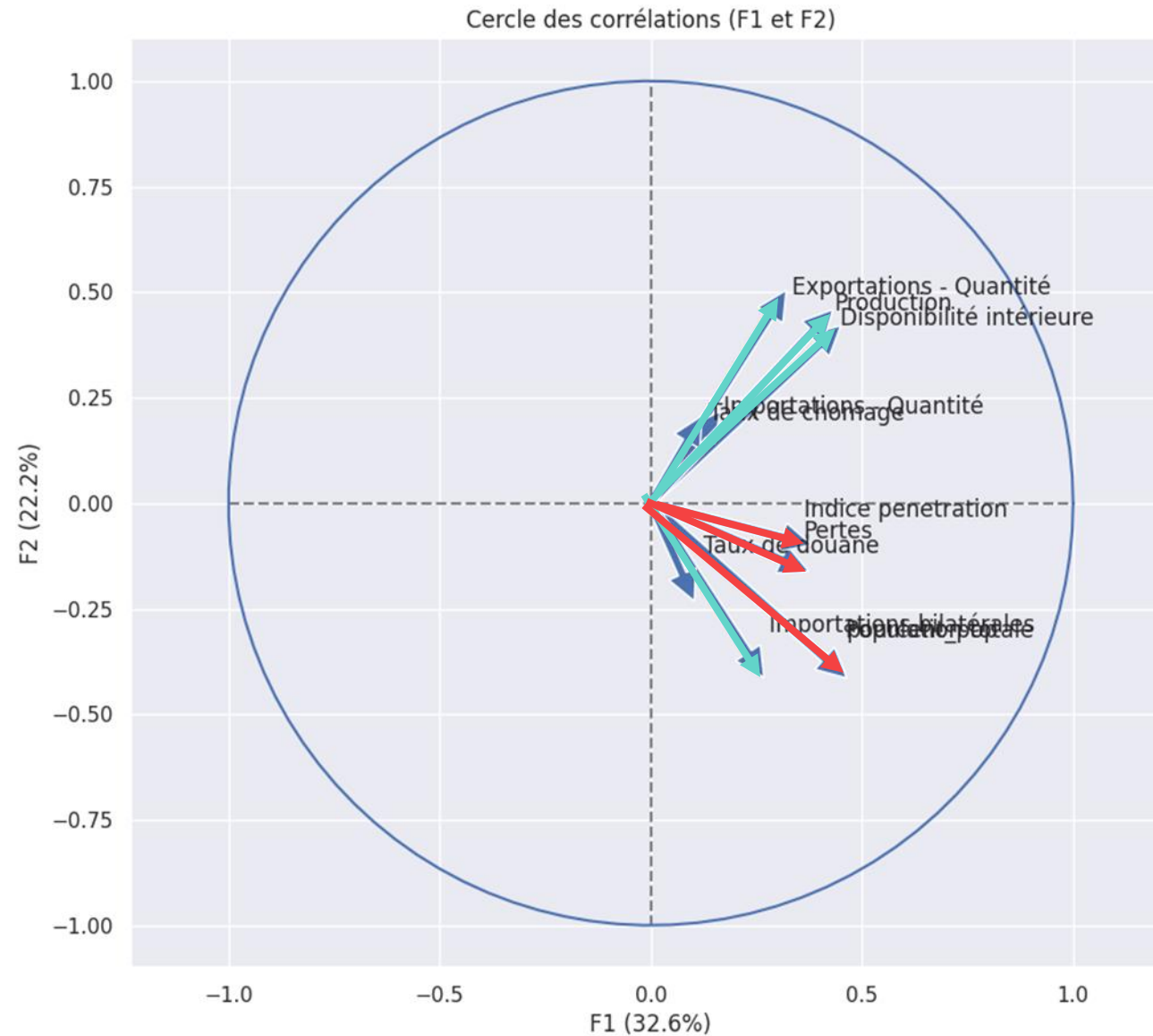
Hypothèse : Demande du pays en importations.

Variable synthétique F4 :

Composition possible : Taux de chômage, taux de douane (anti-corrélé), importations bilatérales (anti-corrélé)

Hypothèse : Situation économique du marché.

Cercle de corrélation des composantes principales F1 et F2



→ Variables les plus corrélées à F1

- Indice de pénétration
- Pertes
- Importations bilatérales

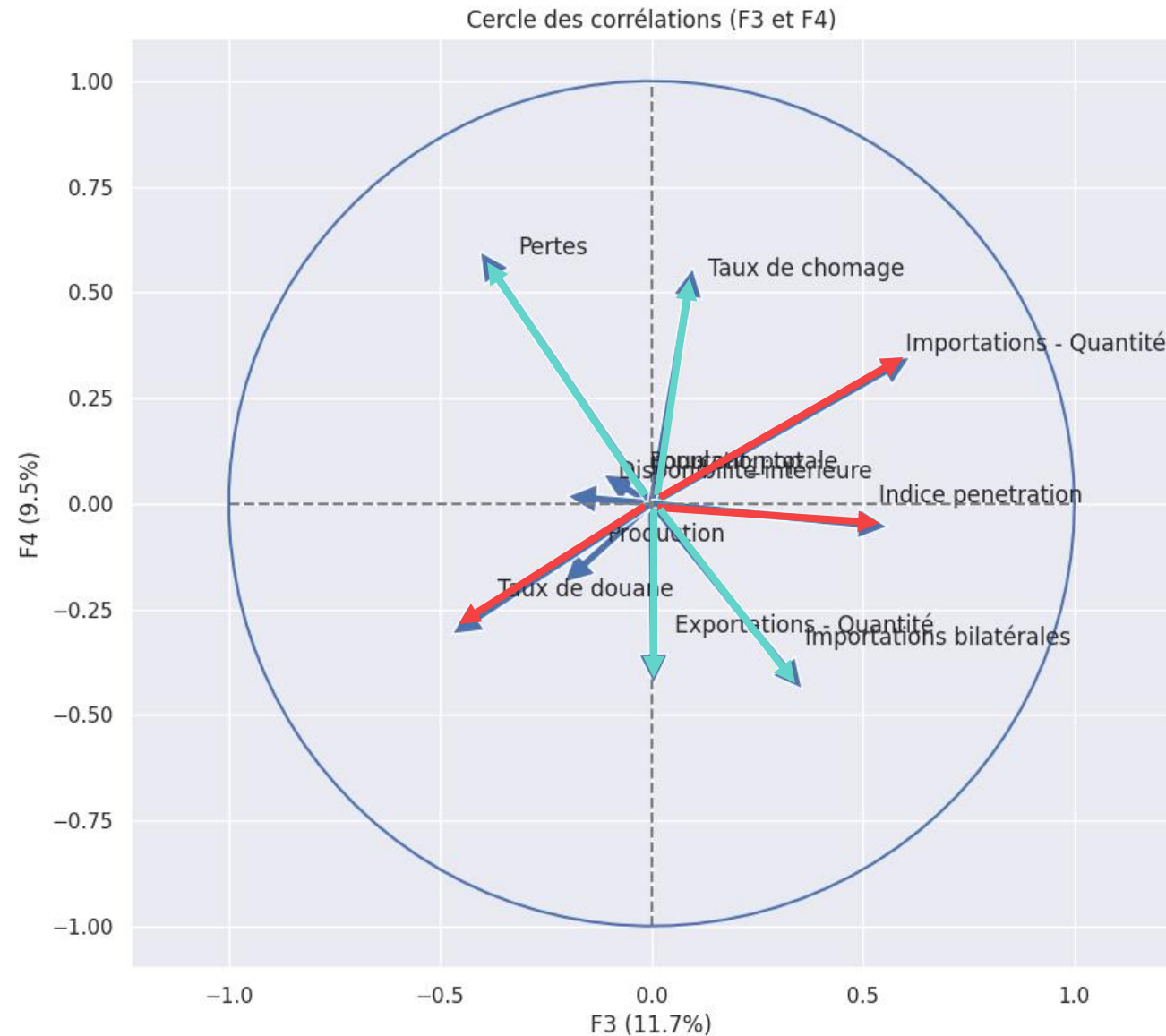
Explication retenue : Echanges commerciaux faciles.

→ Variables les plus corrélées à F2

- Exportations
- Production
- Disponibilité intérieure
- Population totale (anti-corrélé)

Explication retenue : Capacité du pays à satisfaire sa demande en viande de volaille.

Cercle de corrélation des composantes principales F3 et F4



→ Variables les plus corrélées à F3

- Importations
- Indice de pénétration
- Taux de douanes (anti-corrélé)

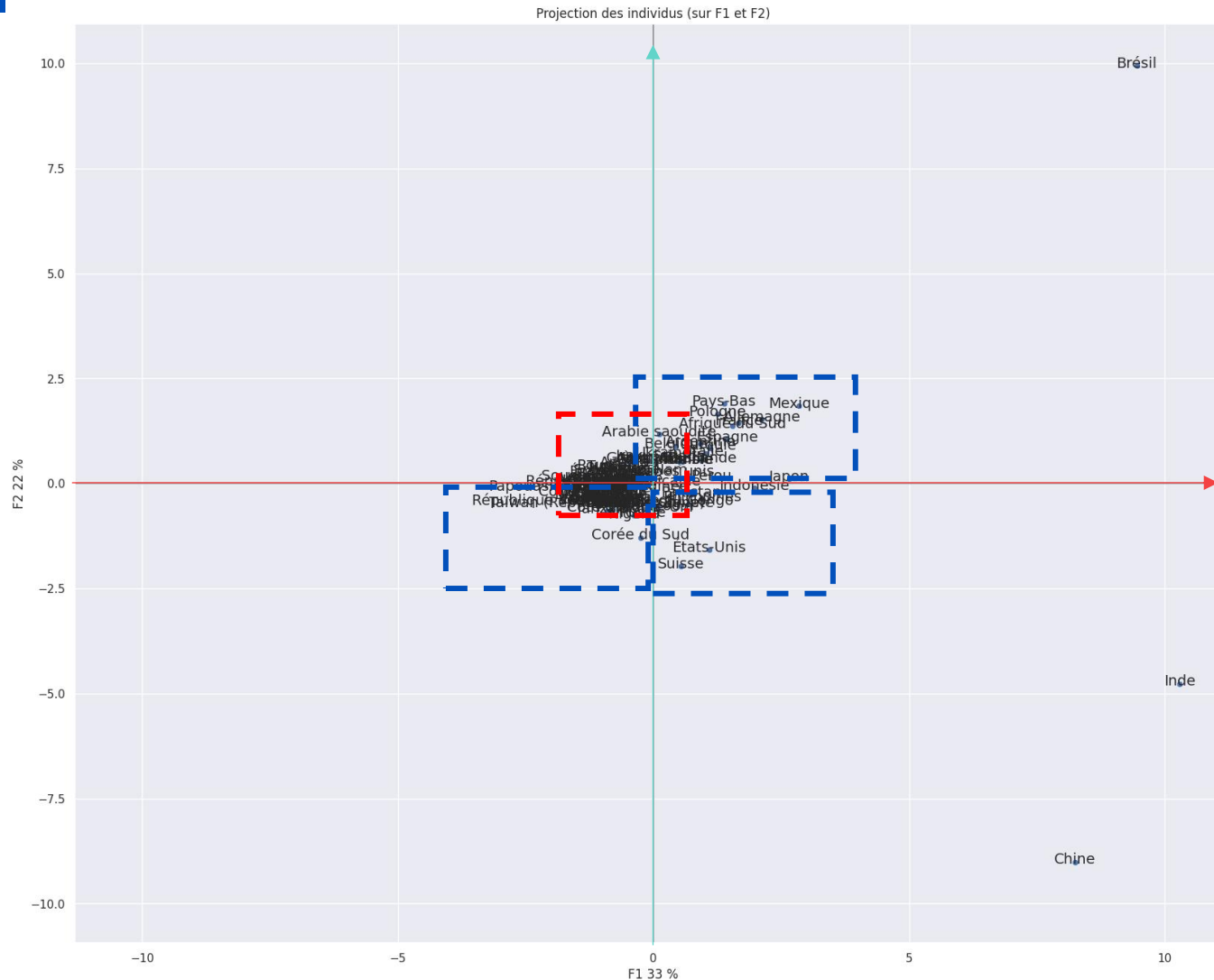
Explication retenue : Demande du pays en importations.

→ Variables les plus corrélées à F4

- Taux de chômage
- Pertes
- Exportations (anti-corrélé)
- Importations bilatérales (anti-corrélé)

Explication retenue : Situation économique du marché.

Projection des composantes principales F1 et F2



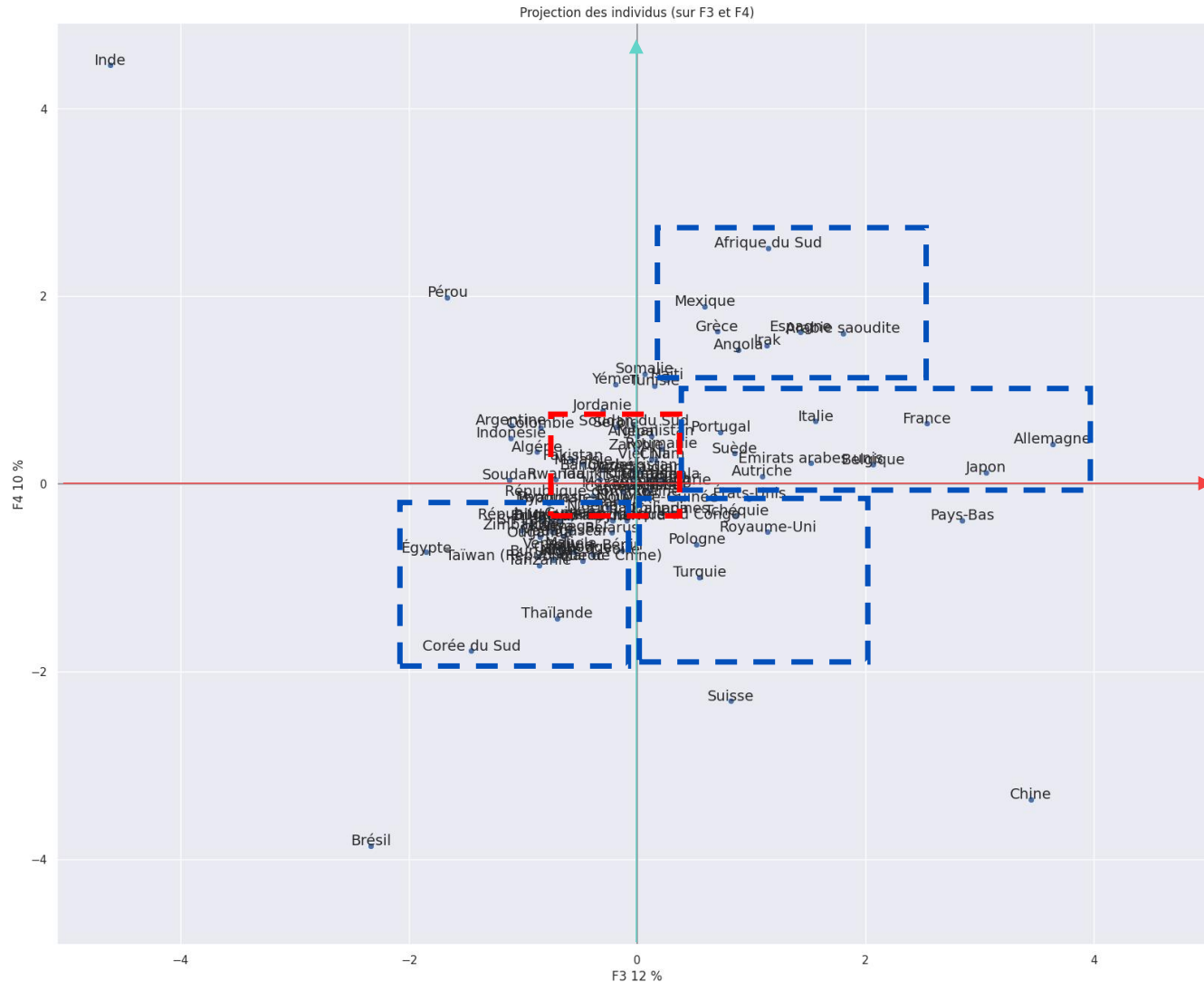
➡ **F1** : Echanges commerciaux faciles.

➡ **F2** : Capacité du pays à satisfaire sa demande en viande de volaille.

Premières interprétations :

- **Inde et Chine** facilitent les échanges commerciaux, et semblent avoir une faible capacité à satisfaire la demande de leur pays.
- Le **Brésil** facilite les échanges commerciaux même s'il est en capacité de satisfaire la demande de son pays.
- On peut distinguer **des tendances de groupement** de pays (bleu) :
 - Un groupe avec des échanges commerciaux faciles et un besoin moyen.
 - Un groupe avec des échanges commerciaux moyens et un besoin plus important.
 - Un groupe avec des échanges commerciaux plus difficiles et un besoin plus important.
- Un certain nombre de pays sont proches du centre 0, **et ne sont pas bien représentés** dans cette projection (rouge)

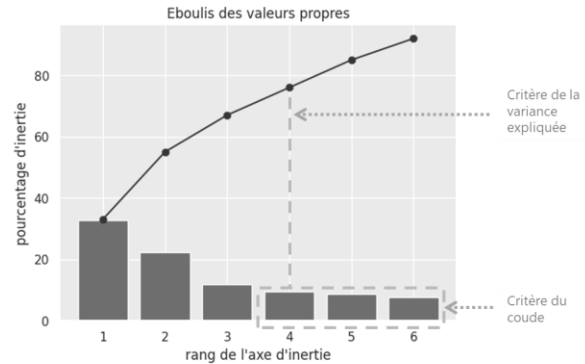
Projection des composantes principales F3 et F4



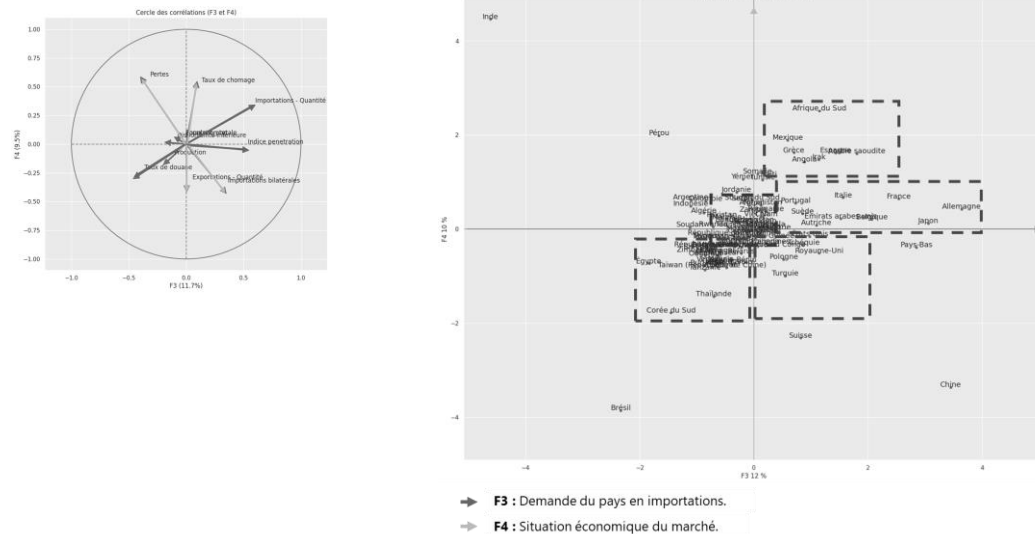
Premières interprétations :

- On note un plus **grand espacement** dans la position des points . **Les regroupements sont plus difficiles** sur cette projection.
- Le **Brésil** est ouvert aux importations même s'il est en capacité de satisfaire la demande de son pays.
- On peut distinguer **des tendances de groupement** de pays (bleu).
 - Deux groupes de pays avec des demandes en importations avérées et des situations économiques favorable. La situation économique est très bonne pour l'un des deux groupes.
 - Un groupe de pays avec une situation économique moins favorable malgré une demande d'importation avérée.
 - Un groupe de pays avec une situation économique moins favorable et des besoins d'importation faibles.
- Un certain nombre de pays sont proche du centre 0, **et ne sont pas bien représentés** dans cette projection (rouge)

ACP en résumé



On a en bleu le pourcentage d'inertie (la variance de chaque nouvelle composante) , et en rouge la variance cumulée.

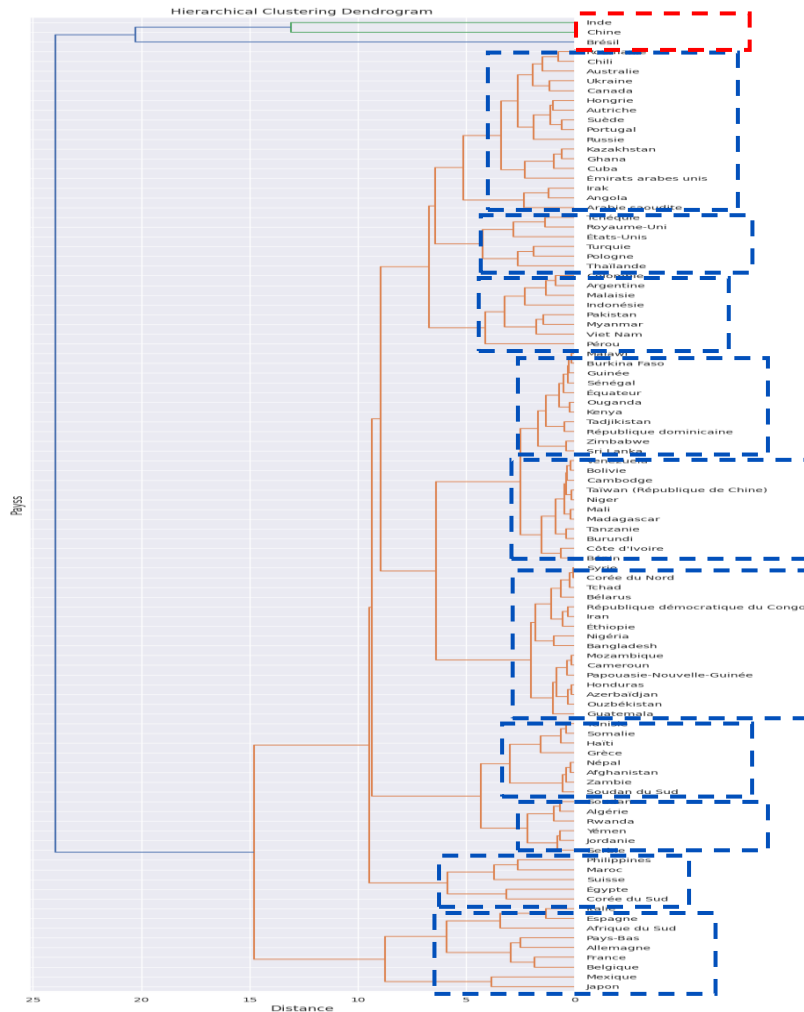


En résumé :

- Nous retenons **4 variables synthétiques**
- Les cercles de corrélations nous permettent d'affiner la composition de ces variables.
- Pour chaque d'elles, l'explication est la suivante :
 - **F1** : Echanges commerciaux faciles.
 - **F2** : Capacité du pays à satisfaire sa demande en viande de volaille.
 - **F3** : Demande du pays en importations.
 - **F4** : Situation économique du marché.
- Ces quatre variables couvrent les **trois premiers axes de l'analyse PESTEL** :
 - **Politique** : F1 et F3
 - **Economique** : F4
 - **Social** : F2
- **Les projections** de pays **selon les composantes** F1, F2,F3 et F4 mettent en évidence **un groupement possible**.

Clustering – comment regrouper les pays ?

Dendrogramme du clustering hiérarchique.



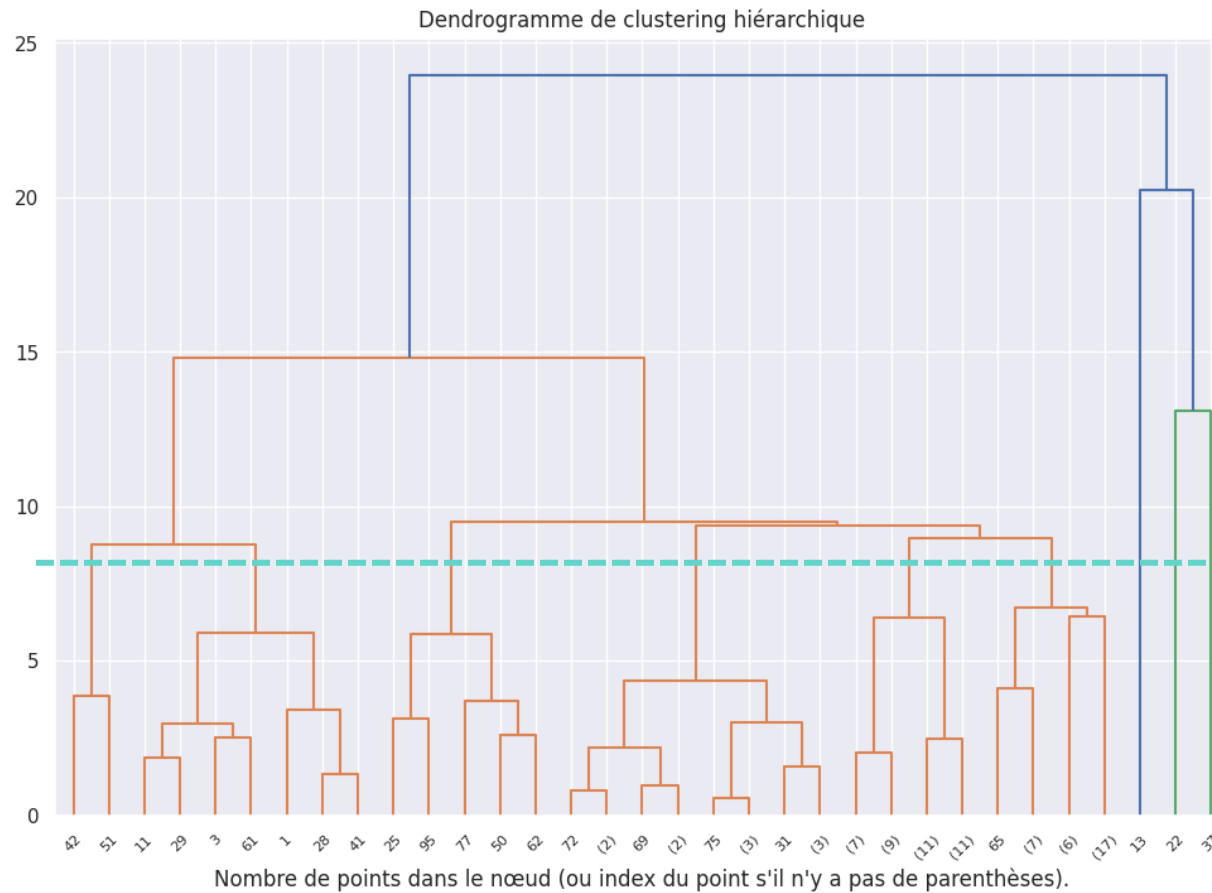
Méthode de clustering Hiérarchique:

- 1 - On regroupe les variables quantitatives dans une matrice
- 2 - on fait un centrage et une réduction des variables
- 3 - On utilise la **fonction Linkage** pour produire une **matrice de liaison** qui représente la **hiérarchie des clusters formés**.
- 4 - On fait une **représentation graphique** de cette hiérarchie avec un **Dendrogramme**.



- On peut écarter trois clusters, formés chacun d'un seul pays (les outliers en rouge).
- Une dizaine de groupes émergent ce qui est encore trop.

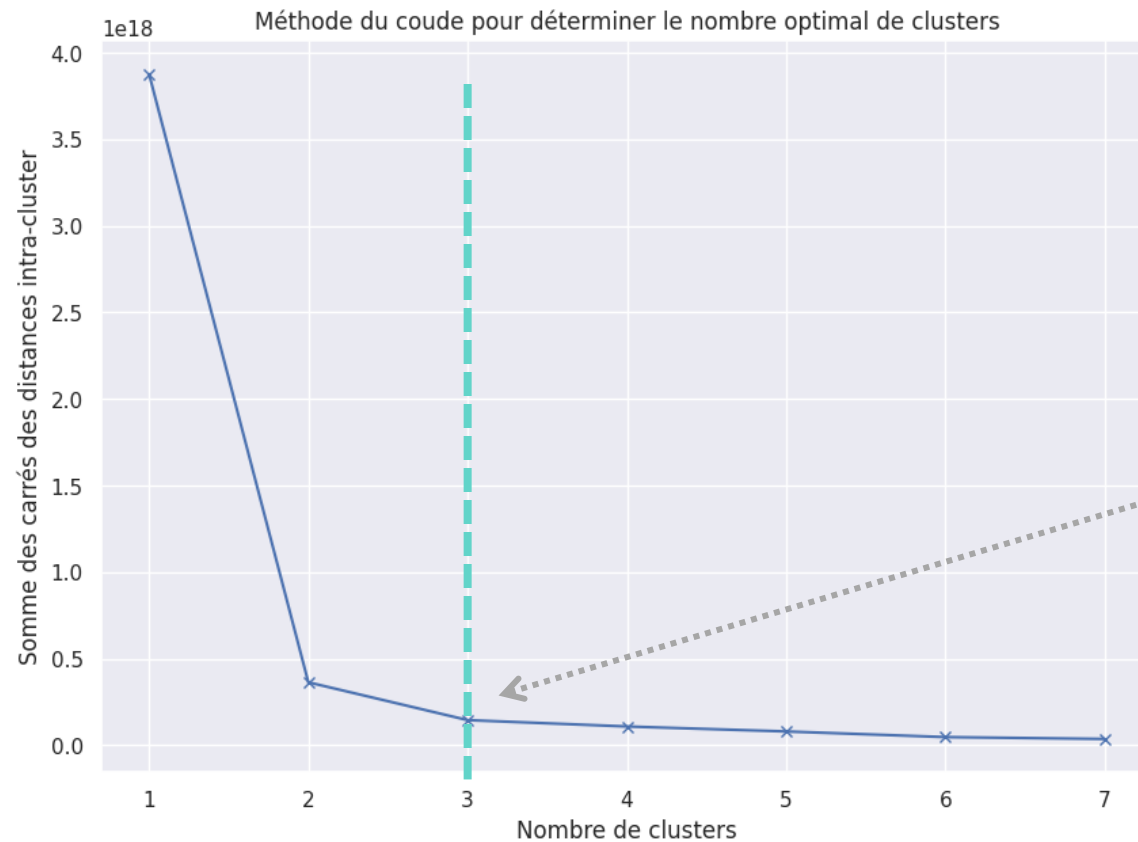
Clustering – comment regrouper les pays ?



On tronque le dendrogramme pour le rendre plus lisible en ne montrant que **les niveaux supérieurs de la hiérarchie**.

Point de coupure (threshold) où les clusters se forment de manière significative, à partir d'une **distance de 10**

Kmeans pour déterminer le nombre de clusters.

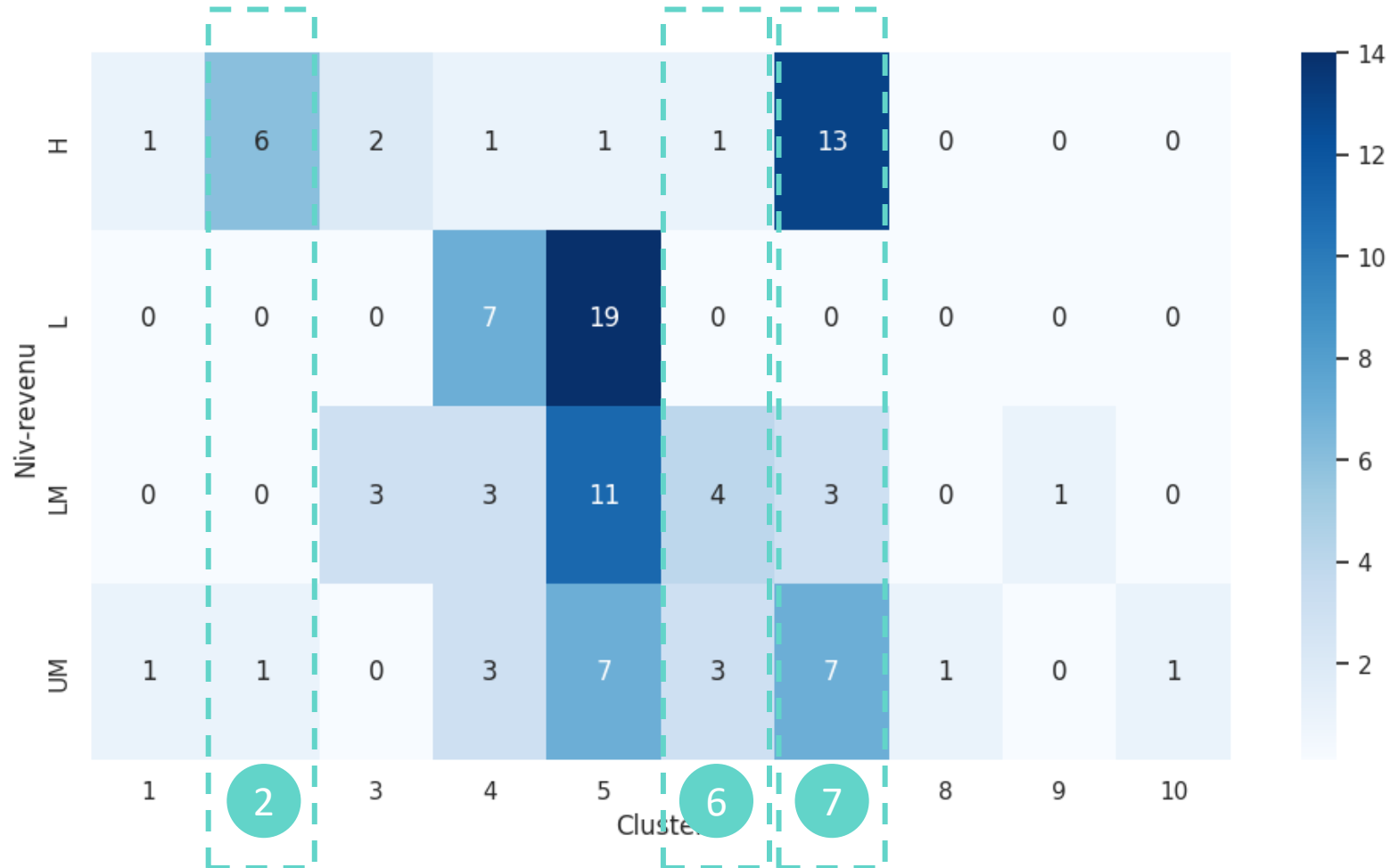


Le point de "coude" est le point où la courbe commence à se stabiliser, c'est-à-dire où l'ajout de plus de clusters ne réduit pas significativement la somme des carrés des distances intra-cluster.



Nous retiendrons **3 clusters..**

Heatmap pour le choix des clusters.



La variable qualitative de niveau de revenu nous permet d'agréger les pays selon le niveau de revenu.

Les cluster 2 et 7 contiennent le plus de pays à fort revenu.

Le cluster 6 est un mix de pays de revenu élevé à moyen mais sans revenus faibles.

Détail des groupes de pays retenus.

Le cluster 2 contient les **7** pays suivants :

- Afrique du Sud (UM)
- Allemagne (H)
- Belgique (H)
- Espagne (H)
- France (H)
- Italie (H)
- Pays-Bas (H)

Le cluster 7 contient les **23** pays suivants :

- Angola (LM)
- Arabie saoudite (H)
- Australie (H)
- Autriche (H)
- Canada (H)
- Chili (H)
- Cuba (UM)
- Ghana (LM)
- Hongrie (H)
- Irak (UM)
- Kazakhstan (UM)
- Pologne (H)
- Portugal (H)
- Roumanie (UM)
- Royaume-Uni (H)
- Russie (UM)
- Suède (H)
- Tchéquie (H)
- Thaïlande (UM)
- Turquie (UM)
- Ukraine (LM)
- Émirats arabes unis (H)
- États-Unis (H)

Le cluster 6 contient les **8** pays suivants :

- Argentine (H)
- Colombie (UM)
- Indonésie (LM)
- Malaisie (UM)
- Myanmar (LM)
- Pakistan (LM)
- Pérou (UM)
- Viet Nam (LM)

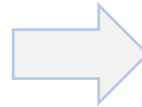
3 Clusters retenus.

38 Pays à étudier.

Conclusion - recommandations

Conclusion :

- Nous avons identifié **3 clusters de pays**.
- Le nombre pays restant **est encore élevé**.



Recommandation :

- Repartir de la liste de pays réduite, ET **ajouter beaucoup plus de variables d'analyses** pour couvrir tous les axes d'analyse PESTEL :
 - Élargir les axes déjà utilisés
 - Couvrir les axes Technologiques, Environnemental et Légal.
- Refaire une ACP et un Clustering.



MERCI

Crédits

Image de rawpixel.com sur Freepik

Image de freepik