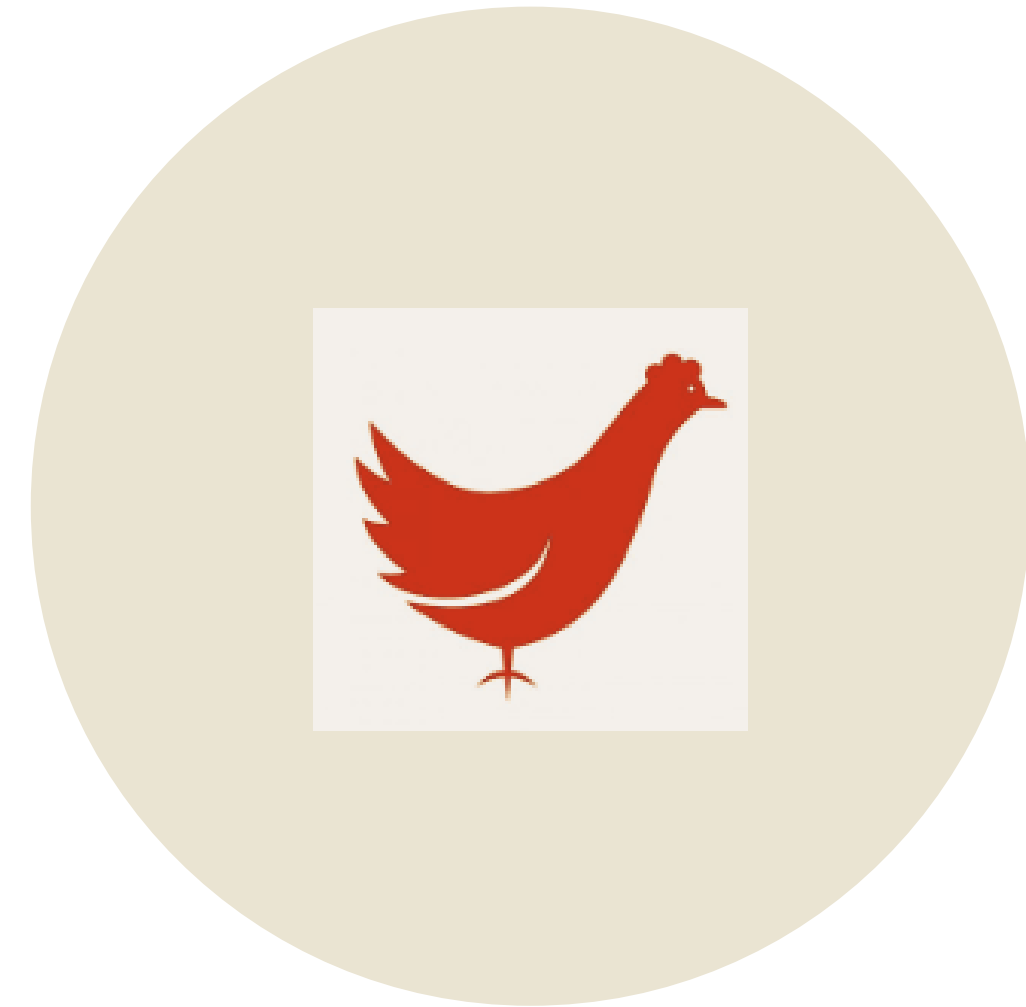
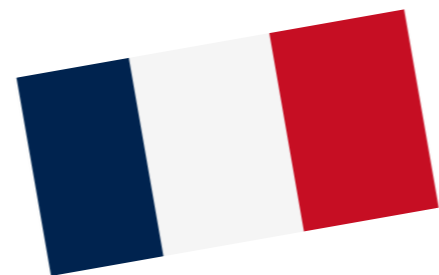




# LA POULE QUI CHANTE



● PRODUISEZ UNE ETUDE DE MARCHE  
AVEC PYTHON



# ENTREPRISE AGROALIMENTAIRE

---

**Stratégie de Développement à l'international**



# ETUDE DE MARCHE

MISSION DATA INTERNATIONALE

---



01

## Analyse des Groupements de pays

Ciblage des pays vers lesquels exporter

02

## Enjeux Economiques et Commerciaux

Opportunités d'augmenter les revenus

Diversification du portefeuille Clients

03

## Axes stratégiques

Recommandations et Conseil

# METHODOLOGIE

## DEMARCHE METHODOLOGIQUE DE NETTOYAGE PUIS D'ANALYSE DE DONNEES

### 01 BASES DE DONNEES PUBLIQUES FAO

### 02 5 FICHIERS CSV

 Disponibilite\_Alimentaire

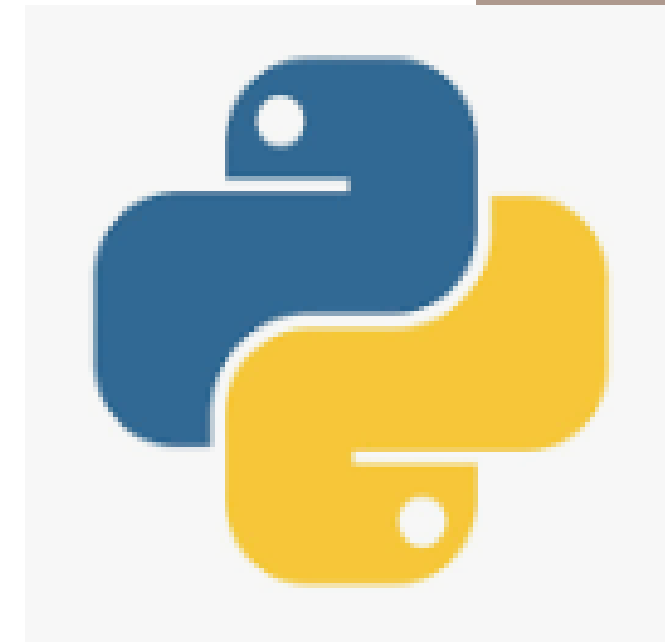
 PIB\_Habitant

 Population\_2000\_2018

 Prix\_Consommation

 Stabilite\_Politique

### 03 LOGICIEL JUPYTER



# SOMMAIRE

1.

**SELECTION DES  
DONNEES**

2.

**PREPARATION  
DES DONNEES**

3.

**ANALYSES**

4.

**RECOMMANDATIONS**

# 1.SELECTION DES DONNEES

---

Indicateurs quantitatifs

- **Population**
- **PIB/habitant en USD**
- **Indicateur Stabilité Politique**
- **Disponibilité de protéines en  
quantité (g/personne/jour)**
- **Exportations - Quantité**
- **Importations - Quantité**
- **Nourriture**
- **Production**

# 2.PREPARATION ET NETTOYAGE DES DONNEES

Faciliter l'analyse et exploration

| 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 | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2001       | 2001  | 1000 personnes | 21606.988 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2002       | 2002  | 1000 personnes | 22600.77  | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2003       | 2003  | 1000 personnes | 23680.871 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2004       | 2004  | 1000 personnes | 24726.684 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2005       | 2005  | 1000 personnes | 25654.277 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2006       | 2006  | 1000 personnes | 26433.049 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2007       | 2007  | 1000 personnes | 27100.536 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2008       | 2008  | 1000 personnes | 27722.276 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2009       | 2009  | 1000 personnes | 28394.813 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2010       | 2010  | 1000 personnes | 29185.507 | X       | Sources internationales sûres | ""   |
| OA           | Séries temporelles annuelles | 2         | Afghanistan | 511          | Population totale | 3010         | Population-Estimations | 2011       | 2011  | 1000 personnes | 30117.413 | X       | Sources internationales sûres | ""   |



Import des données dans Jupyter

```
#Importation du fichier Population_2000_2018
df_population = pd.read_csv('Population_2000_2018.csv')
display(df_population)
```

|   | Code<br>Domaine | Domaine                      | Code<br>zone | Zone        | Code<br>Élément | Élément           | Code<br>Produit | Produit                | Code<br>année | Année | Unité          | Valeur    | Symbole | Description du<br>Symbole     | Note |
|---|-----------------|------------------------------|--------------|-------------|-----------------|-------------------|-----------------|------------------------|---------------|-------|----------------|-----------|---------|-------------------------------|------|
| 0 | OA              | Séries temporelles annuelles | 2            | Afghanistan | 511             | Population totale | 3010            | Population-Estimations | 2000          | 2000  | 1000 personnes | 20779.953 | X       | Sources internationales sûres | NaN  |
| 1 | OA              | Séries temporelles annuelles | 2            | Afghanistan | 511             | Population totale | 3010            | Population-Estimations | 2001          | 2001  | 1000 personnes | 21606.988 | X       | Sources internationales sûres | NaN  |
| 2 | OA              | Séries temporelles annuelles | 2            | Afghanistan | 511             | Population totale | 3010            | Population-Estimations | 2002          | 2002  | 1000 personnes | 22600.770 | X       | Sources internationales sûres | NaN  |
| 3 | OA              | Séries temporelles annuelles | 2            | Afghanistan | 511             | Population totale | 3010            | Population-Estimations | 2003          | 2003  | 1000 personnes | 23680.871 | X       | Sources internationales sûres | NaN  |
| 4 | OA              | Séries temporelles annuelles | 2            | Afghanistan | 511             | Population totale | 3010            | Population-Estimations | 2004          | 2004  | 1000 personnes | 24726.684 | X       | Sources internationales sûres | NaN  |

## 2. PREPARATION ET NETTOYAGE DES DONNEES

Corriger et transformer les données

### Vérification des Valeurs manquantes

```
#Vérification de valeurs nulles  
df_dispo_alimentaire_final.isna().sum()
```

|  |    |
|--|----|
| Élément  |    |
| Disponibilité alimentaire en quantité (kg/personne/an)   | 0  |
| Disponibilité de protéines en quantité (g/personne/jour) | 0  |
| Disponibilité intérieure                                 | 2  |
| Exportations - Quantité                                  | 37 |
| Importations - Quantité                                  | 2  |
| Nourriture   | 2  |
| Production   | 4  |
| Variation de stock                                       | 3  |
| dtype: int64   |    |

### Suppression des colonnes

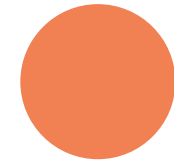
```
#Suppression de colonnes non nécessaires aux analyses  
df_population = df_population.drop(columns=['Année'])  
display(df_population)
```

### Opérations diverses

```
#Renommage de colonnes  
df_population = df_population.rename(columns={'Valeur': 'Population'})
```

```
#Sélection de l'année 2017  
df_population = df_population.loc[df_population["Année"] == 2017]
```

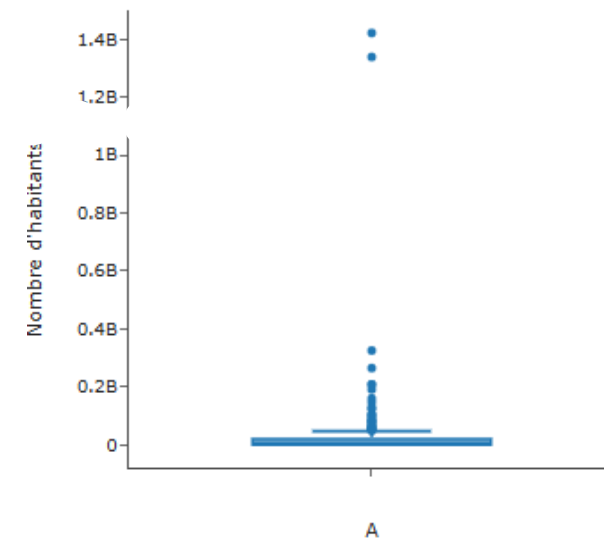




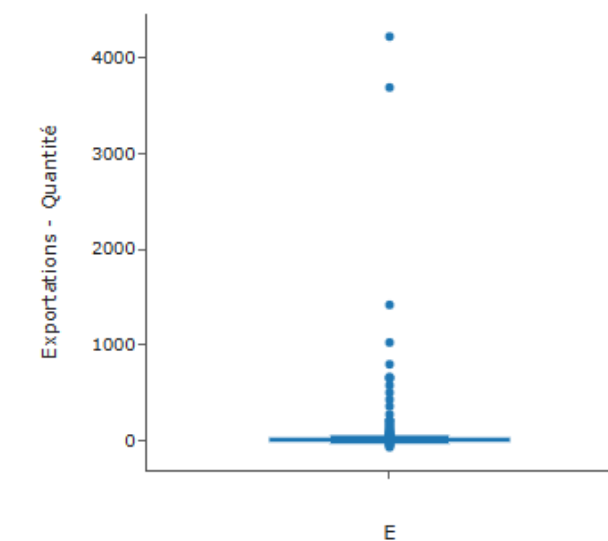
## Traitement des outliers (valeurs extrêmes)

**Chine**  
**Inde**

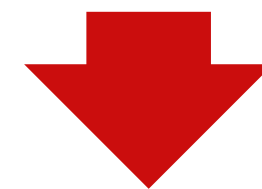
Répartition de la Population des pays



Répartition du volume des Exportations



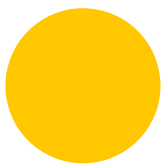
**USA**  
**Bresil**



**Suppression des lignes des 4 Pays**

```
# D'après nos analyses précédentes, il faut enlever les outliers afin de ne pas influencer considérablement nos futures analyses
df_data.drop(df_data[(df_data["Pays"] == "États-Unis d'Amérique")].index, inplace=True)
df_data.drop(df_data[(df_data["Pays"] == "Inde")].index, inplace=True)
df_data.drop(df_data[(df_data["Pays"] == "Brésil")].index, inplace=True)
df_data.drop(df_data[(df_data["Pays"] == "Chine")].index, inplace=True)
```

# 2.PREPARATION ET NETTOYAGE DES DONNEES



Fusion des 5 fichiers en 1 seul



Suppression de lignes manquantes



Nouveau  
Dataframe

|     | Zone                            | Population   | PIB/habitant en USD | Indicateur Stabilité Politique | Disponibilité de protéines en quantité (g/personne/jour) | Exportations - Quantité | Importations - Quantité | Nourriture | Production |
|-----|---------------------------------|--------------|---------------------|--------------------------------|--|-------------------------|-------------------------|------------|------------|
| 1   | Inde                            | 1.338677e+09 | 1937.92             | -0.77                          | 0.75   | 4.00                    | 0.0                     | 2965.0     | 3545.0     |
| 2   | États-Unis d'Amérique           | 3.250848e+08 | 59468.23            | 0.26                           | 19.93  | 3692.00                 | 123.0                   | 18100.0    | 21914.0    |
| 3   | Indonésie                       | 2.646510e+08 | 3839.79             | -0.50                          | 2.42   | 0.00                    | 1.0                     | 1904.0     | 2301.0     |
| 4   | Pakistan                        | 2.079062e+08 | 1558.08             | -2.40                          | 1.97   | 4.00                    | 2.0                     | 1218.0     | 1281.0     |
| 5   | Brésil                          | 2.078338e+08 | 9896.72             | -0.48                          | 15.68  | 4223.00                 | 3.0                     | 9982.0     | 14201.0    |
| ... | ...                             | ...          | ...                 | ...                            | ...  | ...                     | ...                     | ...        | ...        |
| 195 | Grenade                         | 1.108740e+05 | 9309.27             | 0.99                           | 15.50  | -45.70                  | 7.0                     | 5.0        | 1.0        |
| 196 | Saint-Vincent-et-les Grenadines | 1.098270e+05 | 7996.65             | 0.87                           | 25.10  | -70.31                  | 9.0                     | 8.0        | 0.0        |
| 201 | Antigua-et-Barbuda              | 9.542600e+04 | 16110.31            | 0.73                           | 17.77  | 0.00                    | 7.0                     | 5.0        | 0.0        |
| 204 | Dominique                       | 7.145800e+04 | 7395.99             | 1.17                           | 11.52  | 0.00                    | 4.0                     | 3.0        | 0.0        |
| 211 | Saint-Kitts-et-Nevis            | 5.204500e+04 | 22160.60            | 0.63                           | 19.22  | 0.00                    | 4.0                     | 3.0        | 0.0        |

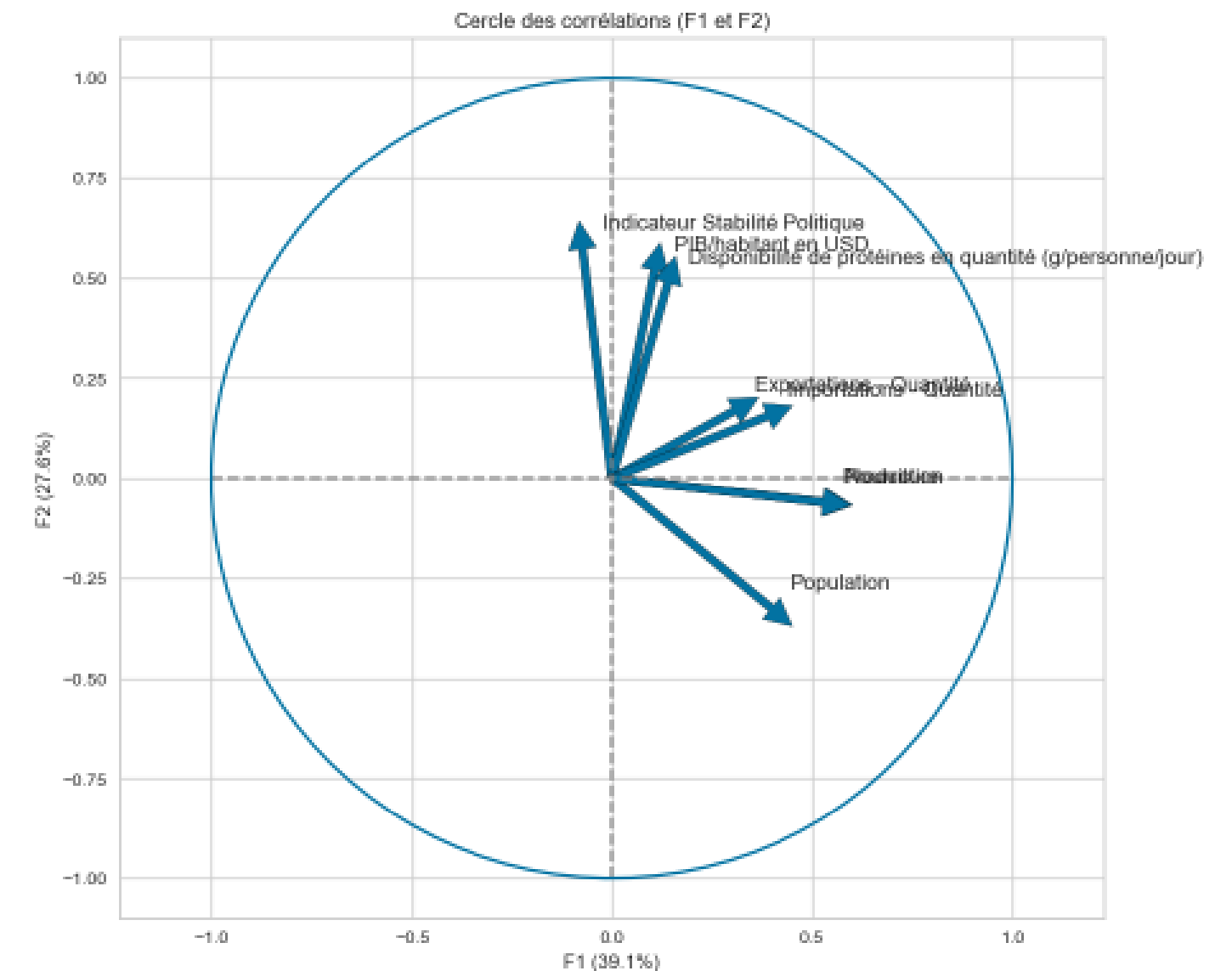
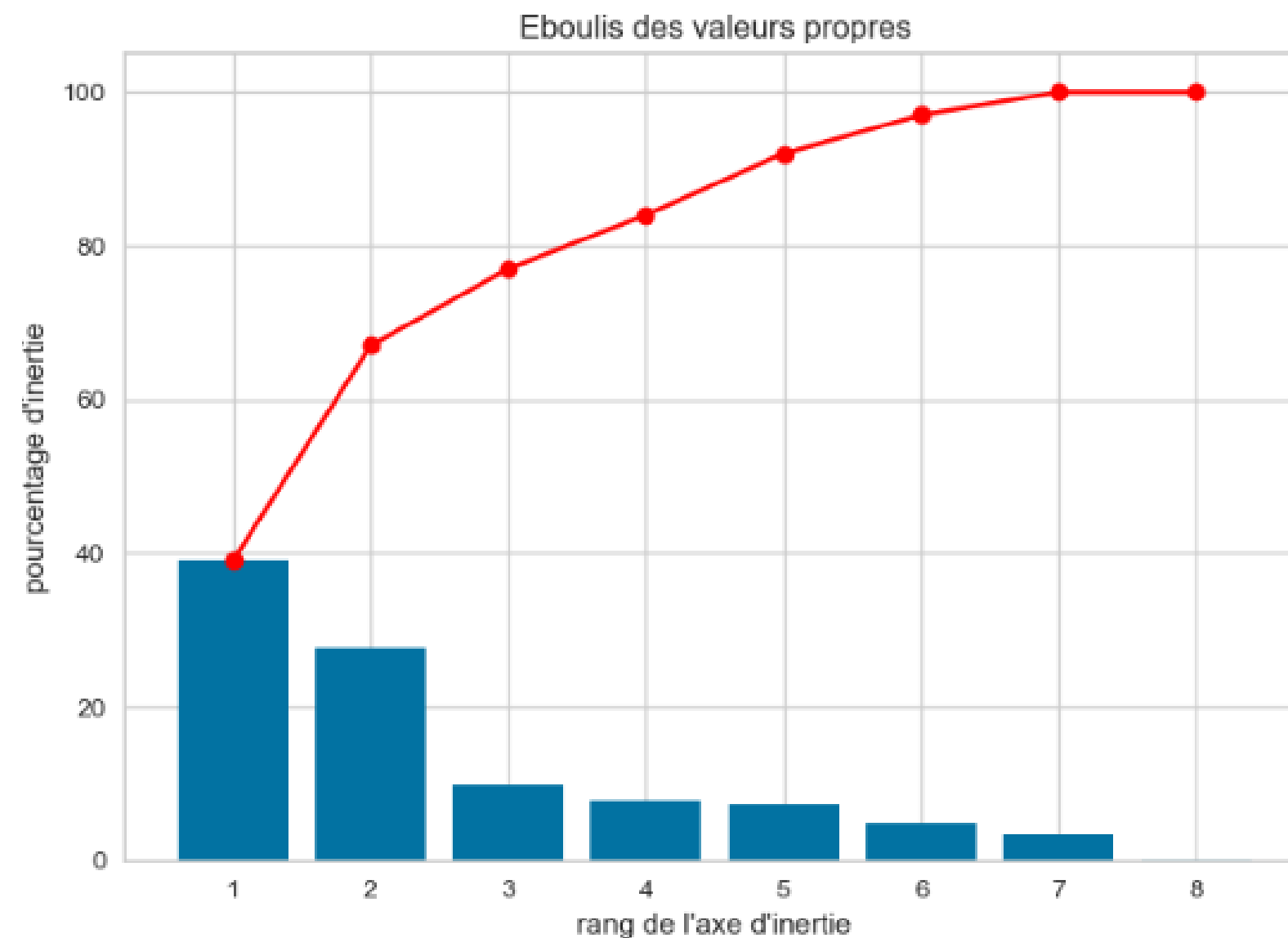
157 lignes  
9 Colonnes

# 3.ANALYSES

## ACP - Analyse en Composantes Principales

Synthétiser les 8 colonnes

Mettre en évidence les liens entre variables



# 3.ANALYSES



## CLUSTERING - Partitionnement

Découpage des Pays en groupes  
avec les mêmes caractéristiques

**Centrage Réduction**

### Scaling

Commençons par scaler les données. On instancie :

```
|: scaler = StandardScaler()
```

On fit :

```
|: scaler.fit(X)
```

```
|: ▾ StandardScaler  
StandardScaler()
```

On transforme :

```
|: X_scaled = scaler.transform(X)  
X_scaled[:155]
```

```
|: array([[ 5.83956824, -0.52574302, -0.48004273, ..., -0.32961364,  
         -0.46272691,  2.6686599 ],  
        [ 4.45442263, -0.64307105, -2.57709413, ..., -0.30328469,  
         -0.45736701,  1.24170215],  
        [ 4.0386461 , -0.62333614, -2.13560962, ..., -0.32961364,  
         -0.46808682, -0.26919429],  
        ...,  
        [-0.61826118,  0.10522057,  0.87752212, ..., -0.32961364,  
         -0.43056748, -0.55038891],  
        [-0.61884624, -0.34287931,  1.36315508, ..., -0.32961364,  
         -0.4466472 , -0.55038891],  
        [-0.61932012,  0.41633312,  0.767151  , ..., -0.32961364,  
         -0.4466472 , -0.55038891]])
```

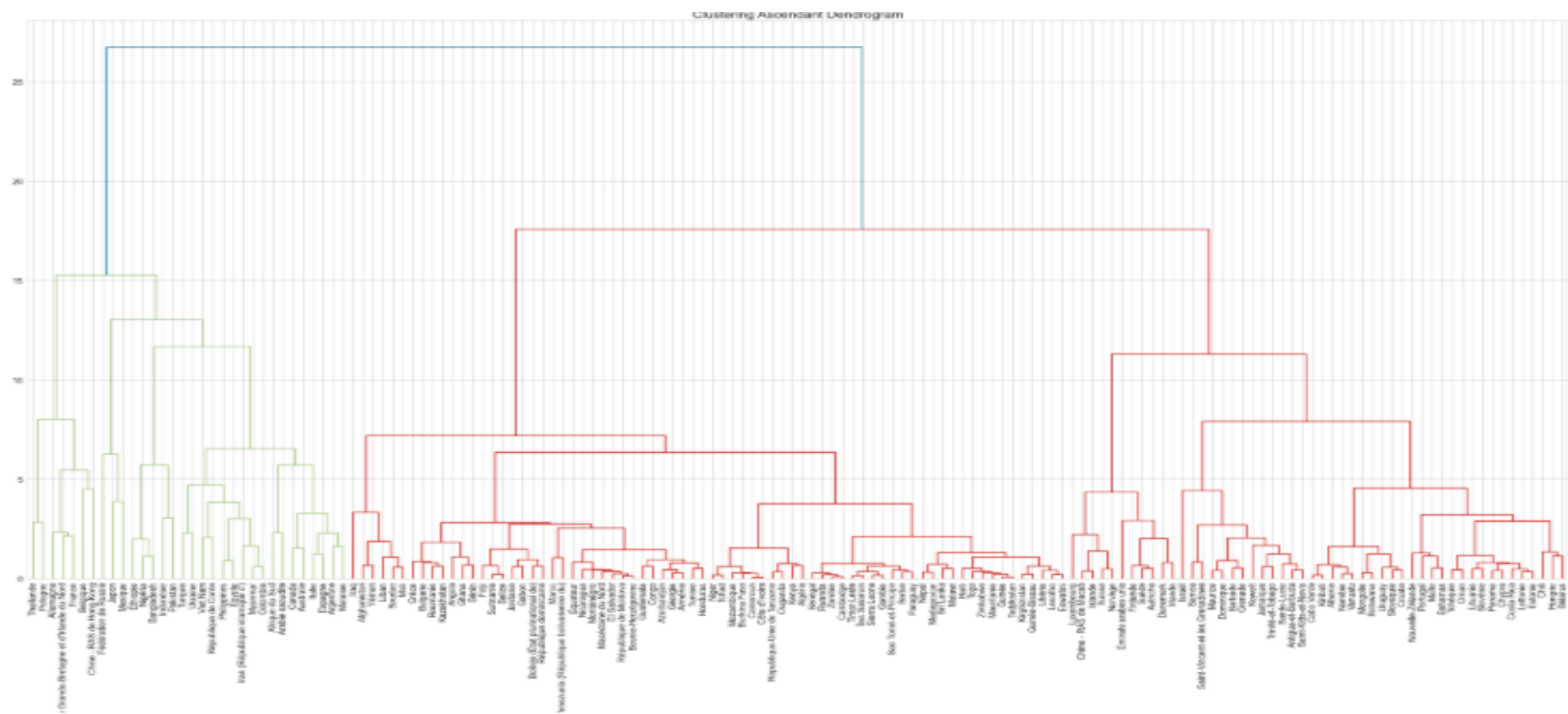
# 3.ANALYSES



## CLUSTERING - Partitionnement

### 1.Classification ascendante hierachique

Regroupement de pays par  
proximité



Dendrogramme

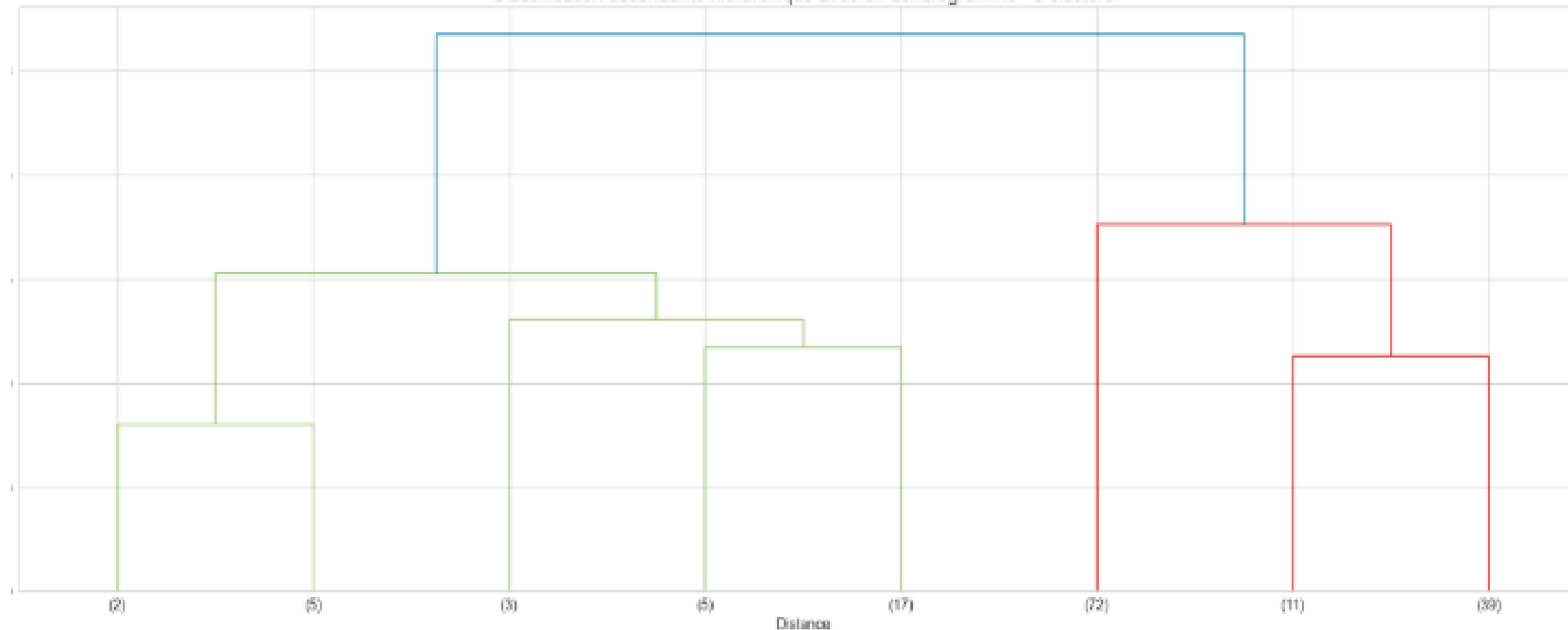
# 3.ANALYSES



## CLUSTERING - Partitionnement

### 1.Classification ascendante hierarchique

Classification ascendante hiérarchique avec un dendrogramme - 8 clusters



### 8 Groupes de Pays

**G1 : 2 pays**

**G2 : 5 pays**

**G3 : 3 pays**

**G4 : 5 pays**

**G5 : 17 pays**

**G6 : 72 pays**

**G7 : 11 pays**

**G8 : 39 pays**

# 3.ANALYSES

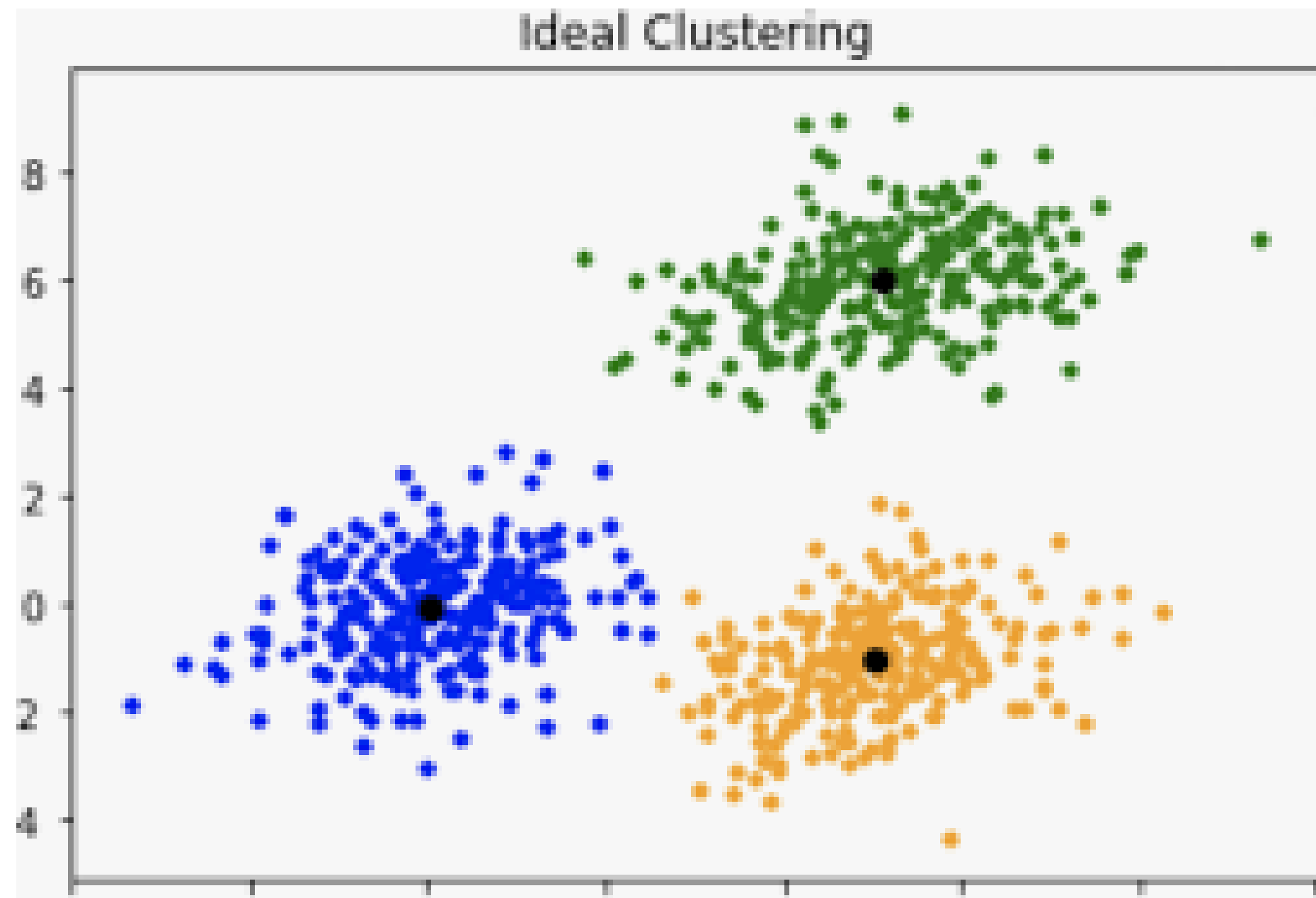
---



## CLUSTERING - Partitionnement

### 2. K Means

Regroupement de Pays grâce  
à des centroïdes



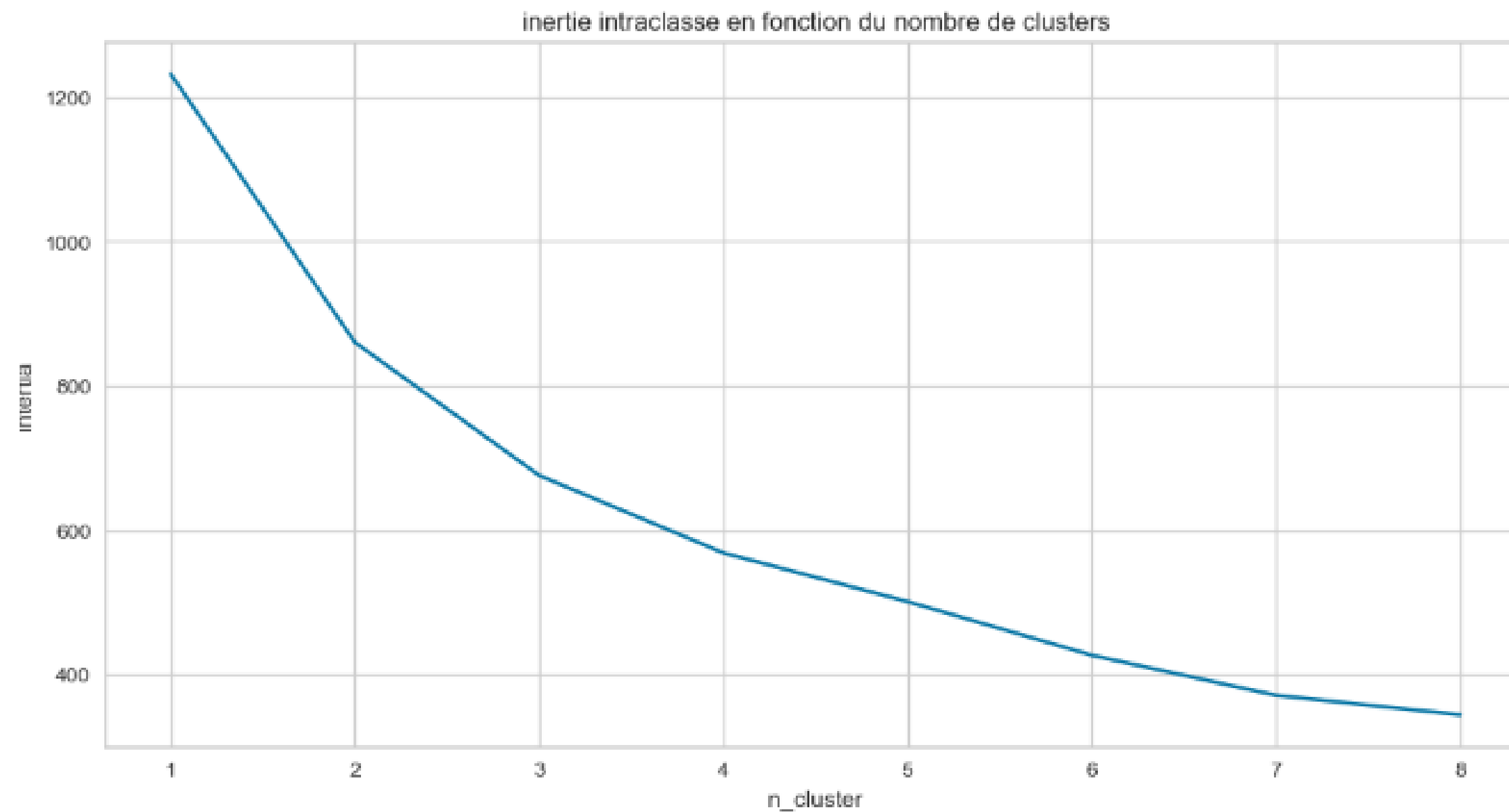
# 3.ANALYSES



## CLUSTERING - Partitionnement

### 2. K Means

Sélection du nombre de groupes  
de Pays





# 3.ANALYSES



## CLUSTERING - Partitionnement

### 2. K Means

#### Attribution du cluster au Pays

```
# Nous devons d'abord ré-entraîner un estimateur !
kmeans_final = KMeans(n_clusters=5)
kmeans_final.fit(df_scaled)
```

| cluster |    |
|---------|----|
| 3       | 88 |
| 0       | 46 |
| 1       | 18 |
| 2       | 6  |
| 4       | 4  |
| ..      | -  |

| cluster |   | Pays                 |
|---------|---|----------------------|
| 0       | 1 | Indonésie            |
| 1       | 1 | Pakistan             |
| 2       | 1 | Nigéria              |
| 3       | 1 | Bangladesh           |
| 4       | 4 | Fédération de Russie |

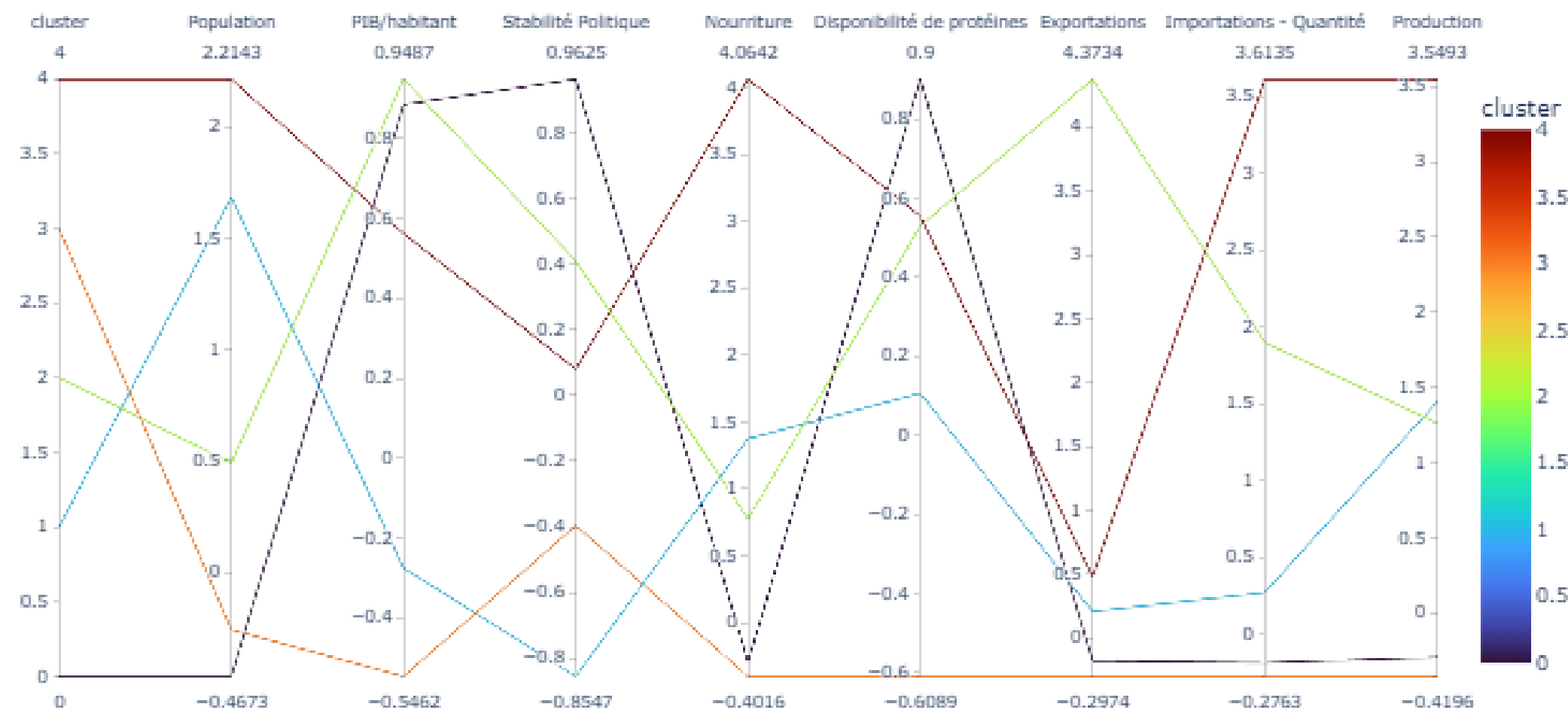
# 3.ANALYSES



## CLUSTERING - Partitionnement

### 2. K Means

Comparaison des groupes avec un diagramme en coordonnées parallèles



PIB / Stabilité Politique : G0,G2

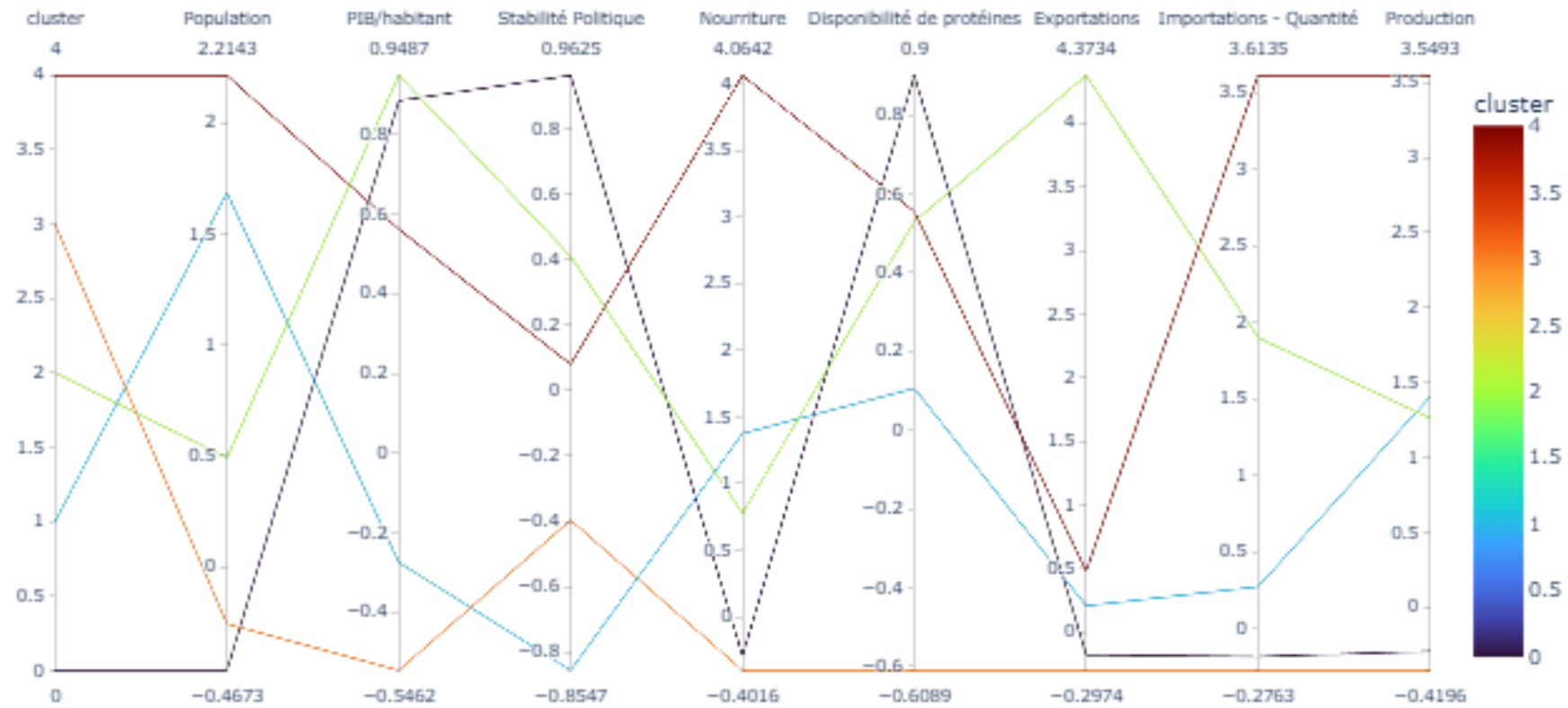
Nourriture / Disponibilité : G4,G0,G2

Importations : G4,G2

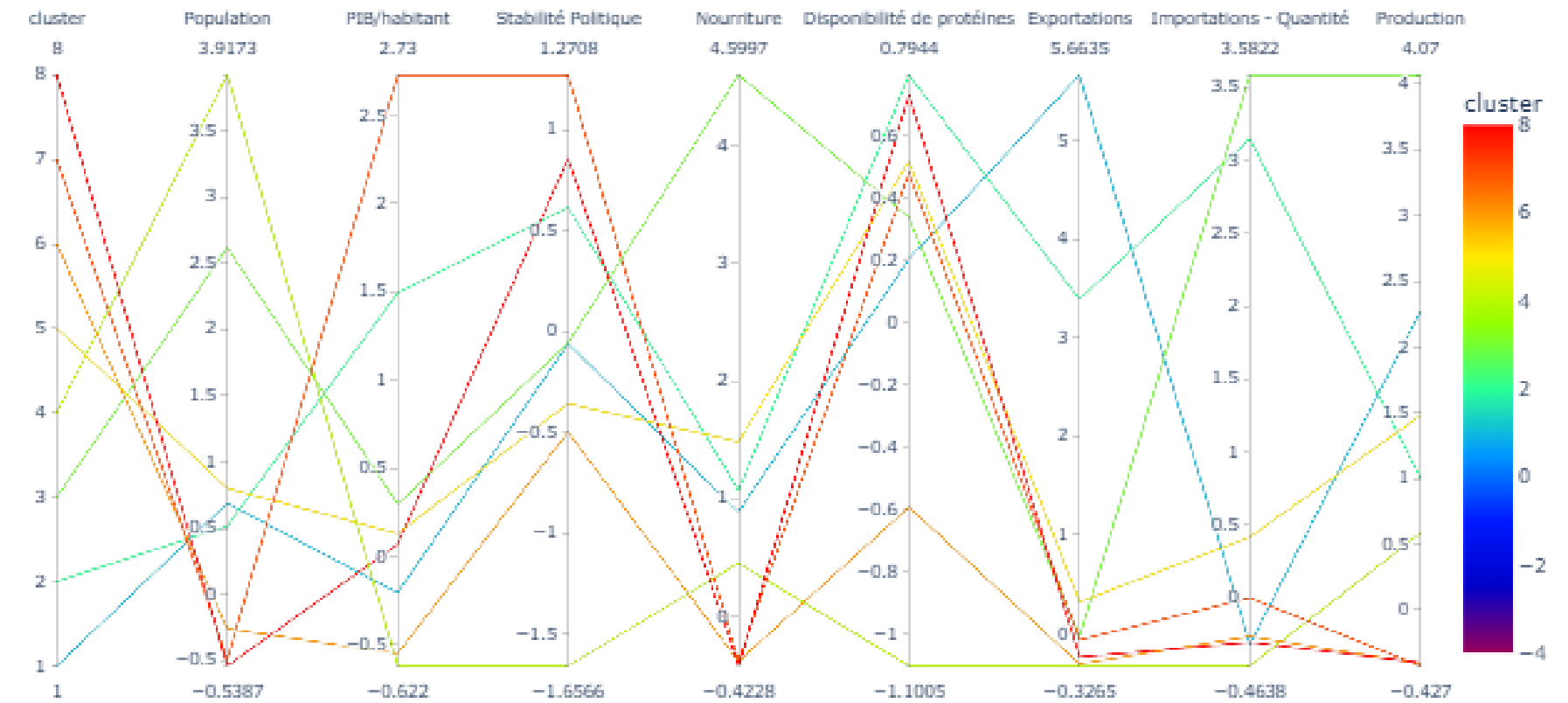
Production : G4

Population : G1,G4

# 4.RECOMMENDATIONS

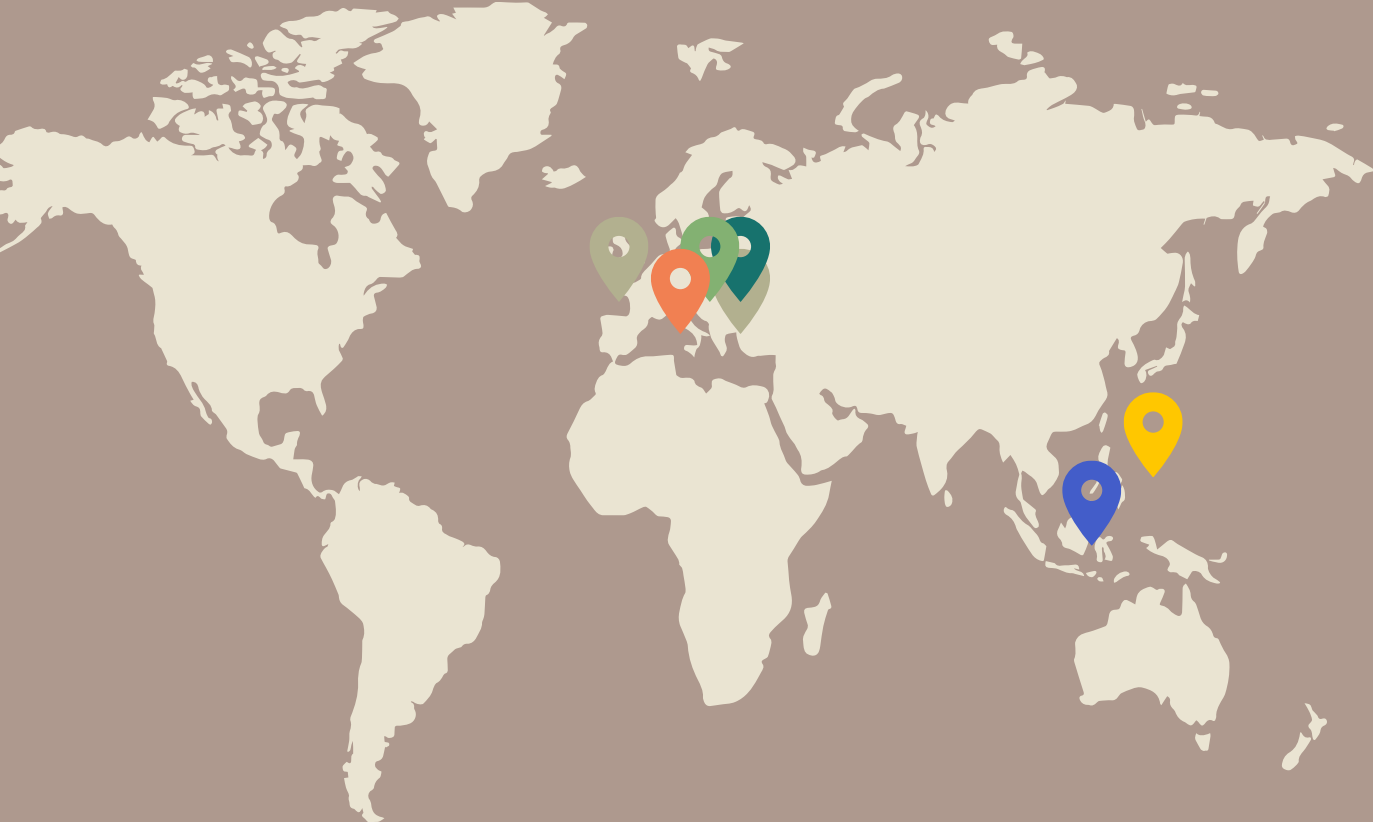


**KMEANS**



**Classification ascendante hierarchique**

# PAYS CIBLES



- Allemagne
- Thaïlande
- Royaume-Uni de Grande-Bretagne et d'Irlande du Nord
- Pologne
- Belgique
- Chine - RAS de Hong-Kong

## Des pays à majorité européens

|    | Pays  | Population | PIB/habitant en USD | Indicateur Stabilité Politique | Disponibilité de protéines en quantité (g/personne/jour) | Exportations - Quantité | Importations - Quantité | Nourriture | Production |
|----|---|------------|---------------------|--------------------------------|--|-------------------------|-------------------------|------------|------------|
| 11 | Allemagne   | 82658409.0 | 44670.22            | 0.57                           | 7.96   | 646.0                   | 842.0                   | 1609.0     | 1514.0     |
| 14 | Thaïlande   | 69209810.0 | 6436.79             | -0.75                          | 4.35   | 796.0                   | 2.0                     | 896.0      | 1676.0     |
| 15 | Royaume-Uni de Grande-Bretagne et d'Irlande du... | 66727461.0 | 40568.47            | 0.38                           | 13.77  | 359.0                   | 779.0                   | 2131.0     | 1814.0     |
| 30 | Pologne   | 37953180.0 | 13615.44            | 0.51                           | 12.14  | 1025.0                  | 55.0                    | 1150.0     | 2351.0     |
| 65 | Belgique  | 11419748.0 | 44162.26            | 0.42                           | 4.57   | 656.0                   | 338.0                   | 144.0      | 463.0      |
| 87 | Chine - RAS de Hong-Kong                          | 7306322.0  | 45737.48            | 0.82                           | 22.26  | 663.0                   | 907.0                   | 391.0      | 24.0       |

## Des pays à forte stabilité économique et politique

275 274 930 d'habitants  
32 531\$ de Pib/habitant contre 14 064\$ pour le monde  
Environnement prévisible qui attire les investissements  
Importations conséquentes donc demande présente  
Cas Thaïlande

# MERCI

● POUR VOTRE ATTENTION

