



ETUDE DE MARCHE CLUSTERING

PROJET 9

Jeu de données : 18 variables

- Volume de volaille :

- ⇒ Disponibilité alimentaire (Kcal/personne/jour)
- ⇒ Disponibilité intérieure
- ⇒ Exportations – Quantité
- ⇒ Importations – Quantité
- ⇒ Production

- Démographie / Economie / Politique :

- ⇒ Population
- ⇒ Taux de variation de la population entre 2007 et 2017
- ⇒ PIB par habitant
- ⇒ Indice de stabilité politique
- ⇒ Taux de variation du taux de change entre 2007 et 2017

- Représentation et dynamique de la volaille :

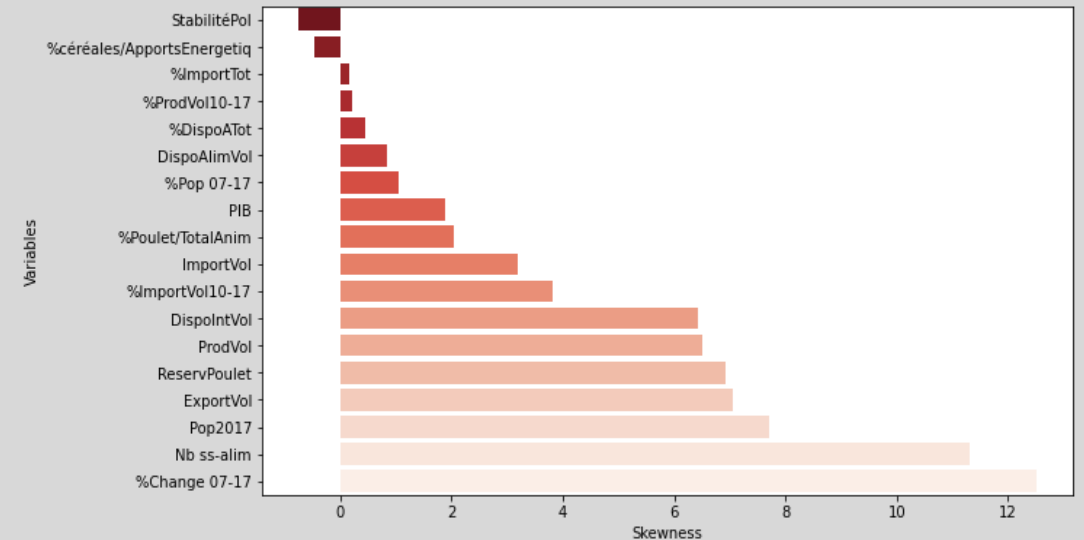
- ⇒ Part de la volaille dans la disponibilité alimentaire en viande
- ⇒ Part de la volaille dans les importations de viande
- ⇒ Taux de variation de la production de volaille entre 2010 et 2017
- ⇒ Taux de variation des importations de volaille entre 2010 et 2017

- Capacité de production et sécurité alimentaire :

- ⇒ Part des poulets dans le total des animaux
- ⇒ Réserves en poulet (en Unités Gros Bétail)
- ⇒ Nombre de personnes sous-alimentées (en millions)
- ⇒ Part des céréales, racines et tubercules dans les apports énergétiques alimentaires

Normalisation du dataset

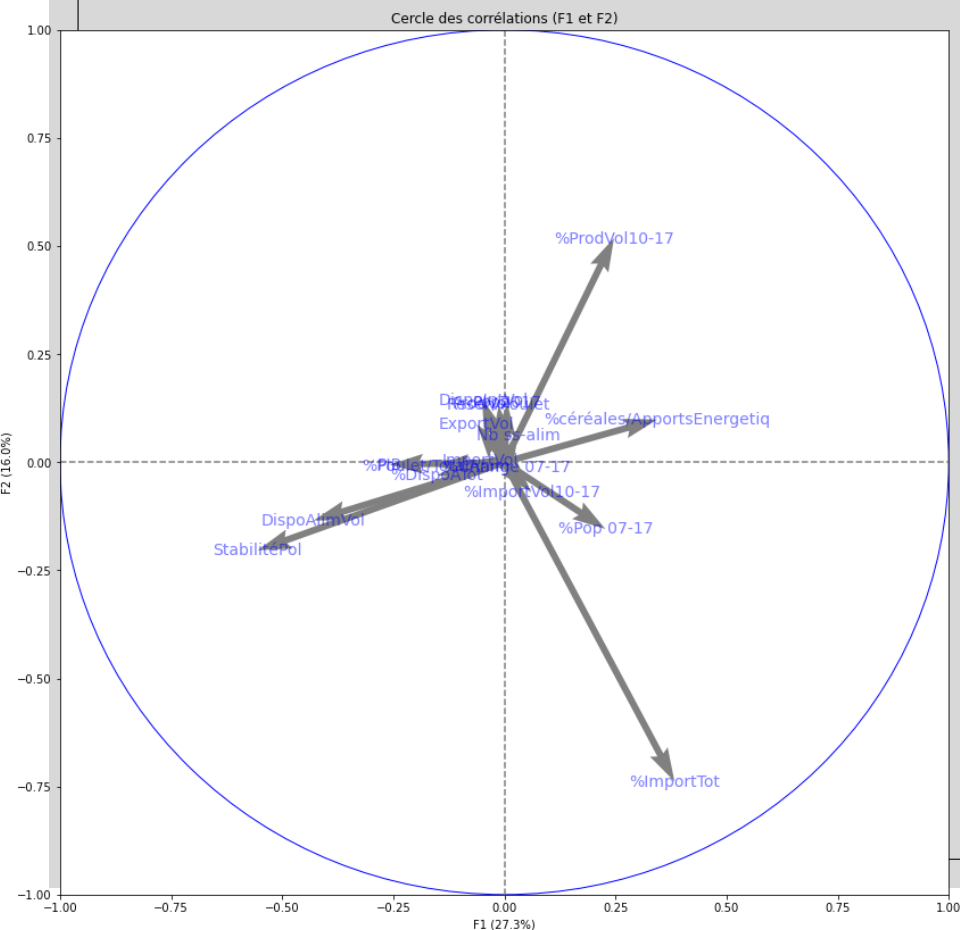
- Pour la plupart des variables :
 - ⇒ skewness > 0 : distribution des données étalée à droite et ne suit pas une loi normale,
 - ⇒ coefficient de variation > 100% : dispersion
- Mise à l'échelle des données :
 - ⇒ Lorsque distribution non normale, StandardScaler n'est pas recommandé,
 - ⇒ présence à la fois de valeurs positives et négatives dans certaines colonnes,
 - ⇒ Choix de **MaxAbsScaler**



Méthode sensible à la présence d'outliers

1. Analyse en composantes principales

14 composantes permettent d'expliquer 99% de la variance totale

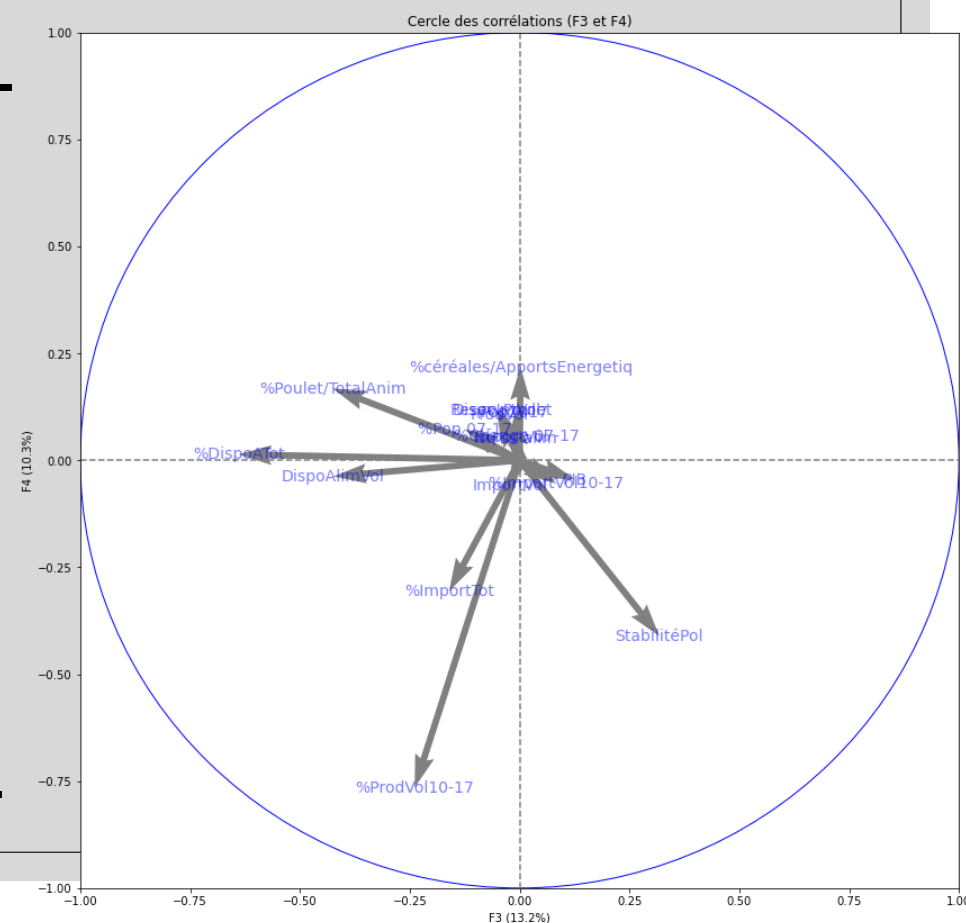


F1 : + Développement –

**F2 : - Effort
d'autosuffisance +**

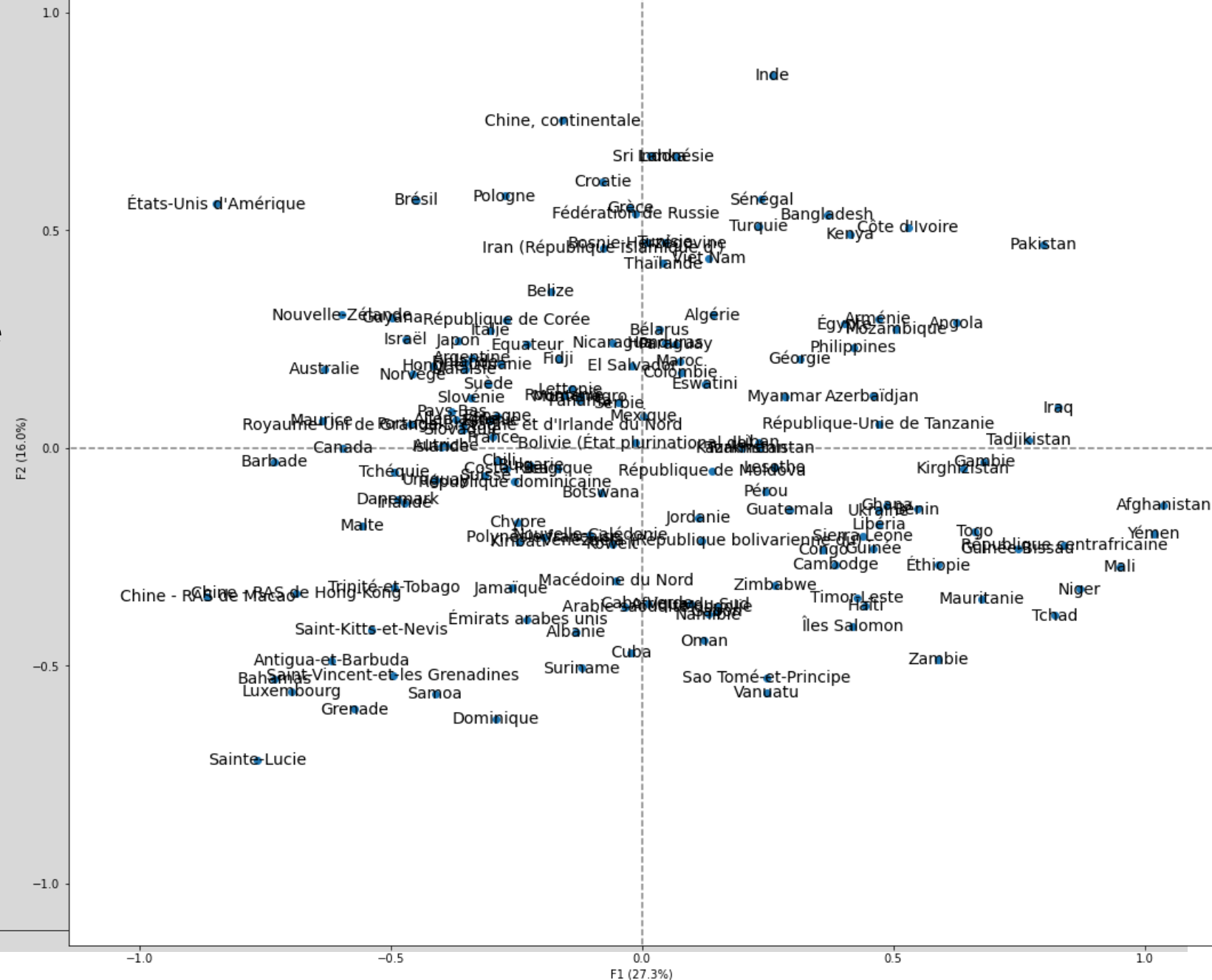
**F3 : + Spécialisation sur
la volaille –**

**F4 : + Dynamique de
production de volaille –**



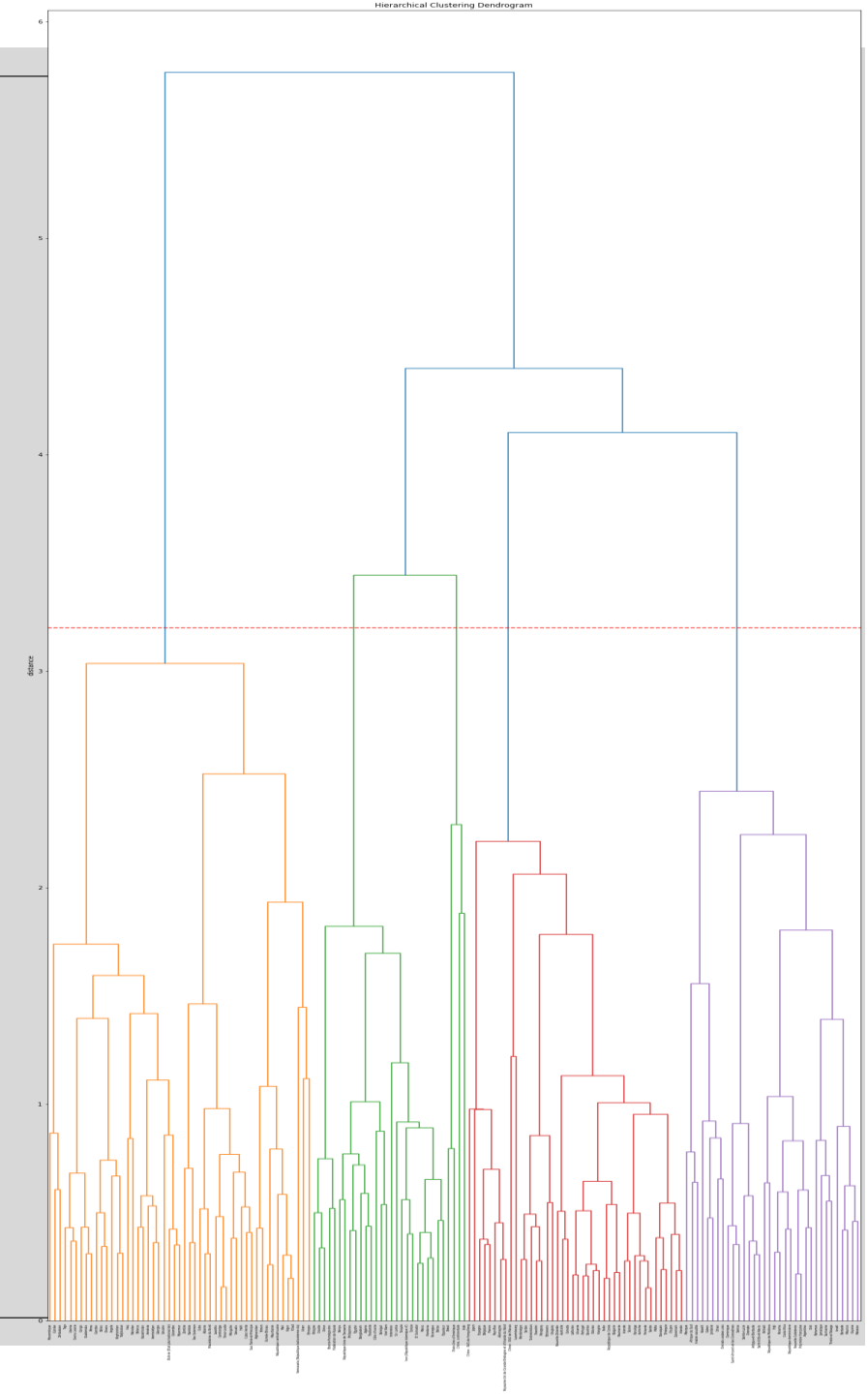
Projection des individus (sur F1 et F2)

- l'Afghanistan, le Mali ou le Yémen feraient partie des pays les moins « développés » (F1)
- l'effort d'autosuffisance serait important pour l'Inde, la Chine ou la Croatie (F2)



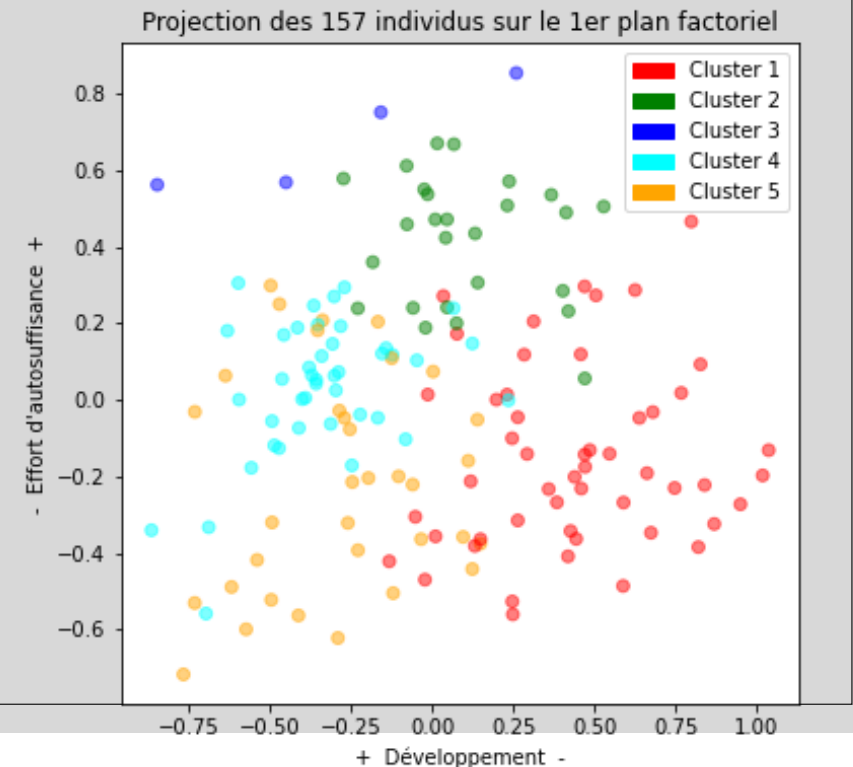
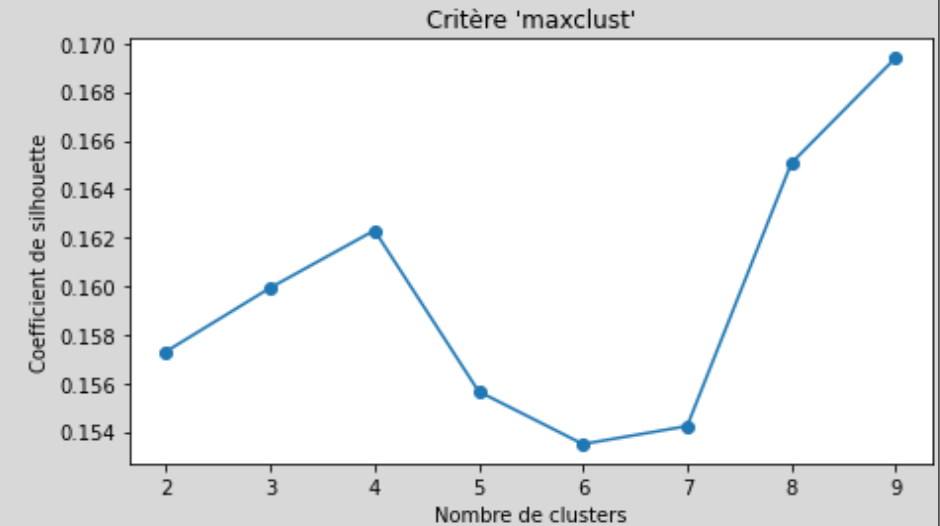
2. Classification ascendante hiérarchique

- Plusieurs choix possibles de méthode de lien :
 - ⇒ choix de la méthode de Ward (malgré un coefficient de corrélation cophénétique moins élevé que les autres méthodes)
 - ⇒ méthode qui cherche à minimiser l'augmentation de l'inertie intraclasse
- Le dendrogramme : arbre qui découpe les différents clusters



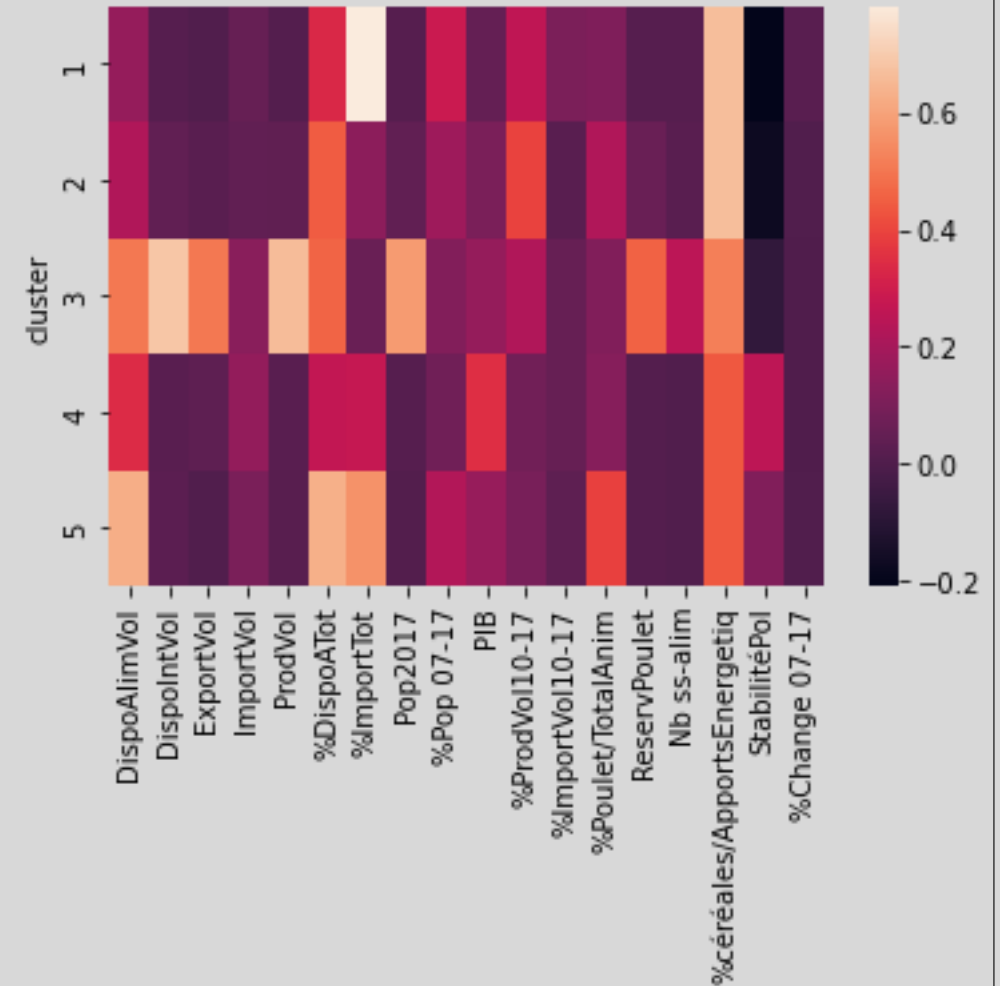
CAH : choix des clusters et projection

- Un score silhouette :
 - ⇒ + élevé pour la création de 9 clusters,
 - ⇒ relativement similaire peu importe le nombre de clusters (compris entre 0.15 et 0.17),
 - ⇒ plutôt proches de 0 (peu d'étanchéité entre les clusters)
- Choix de 5 clusters :
 - ⇒ Dont 1 qui regroupe 4 pays (Brésil, Etats-Unis, Inde, Chine) – peuvent être considérés comme outliers



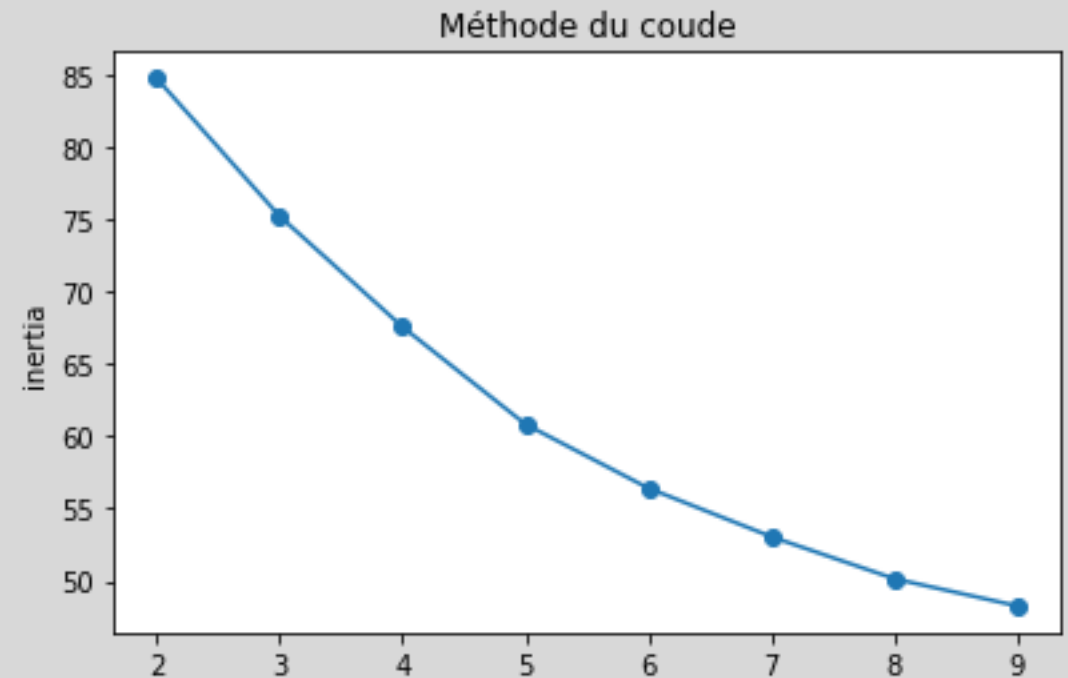
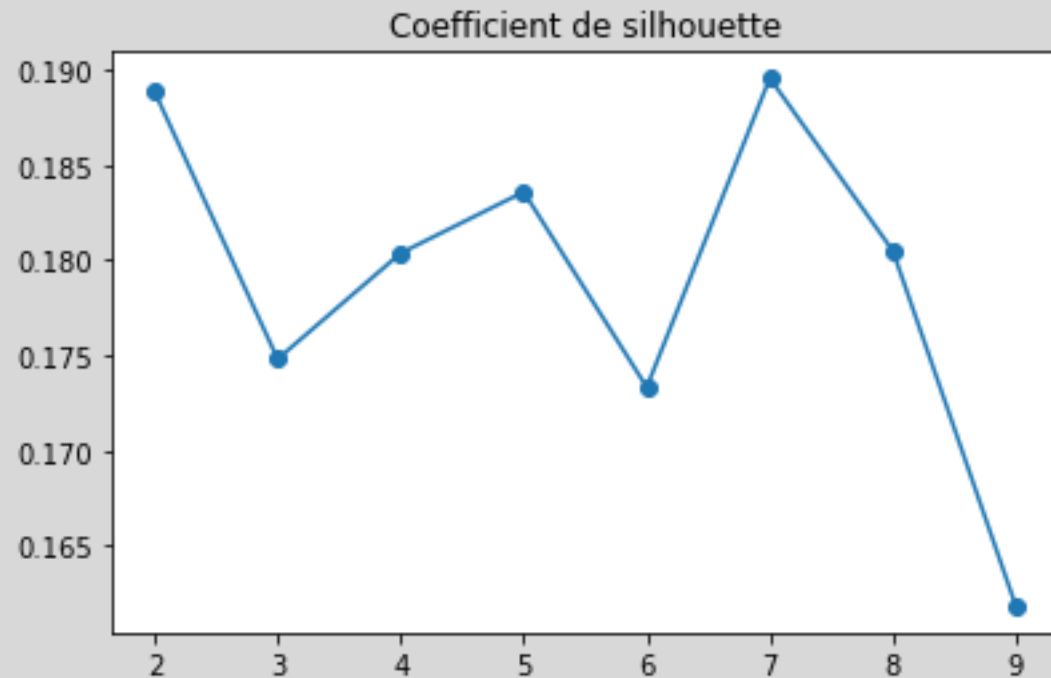
Analyse des caractéristiques des clusters

- Cluster 1 :
 - ⇒ Carence en viande,
 - ⇒ Importations en viande centrées sur la volaille
 - ⇒ Instabilité économique et politique
- Cluster 3 :
 - ⇒ Forte capacité de production
 - ⇒ Faible niveau d'importation de volaille en comparaison aux autres viandes
- Cluster 5 :
 - ⇒ Forte disponibilité en volaille
 - ⇒ Dispo. générée par la capacité de production et/ou les importations



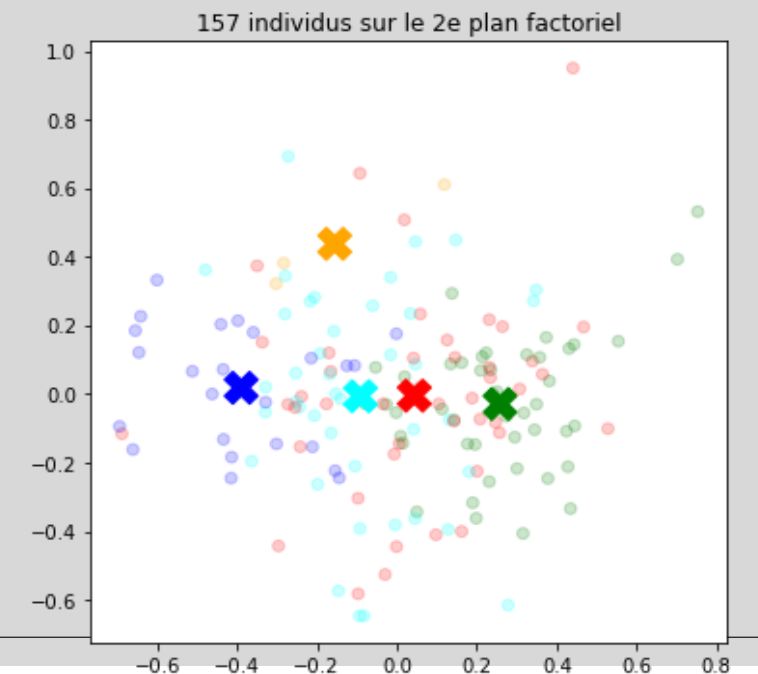
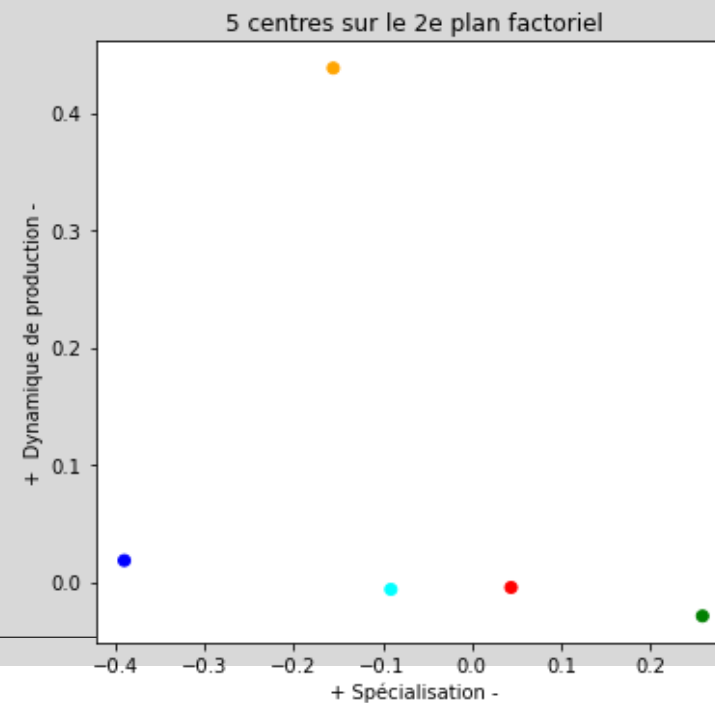
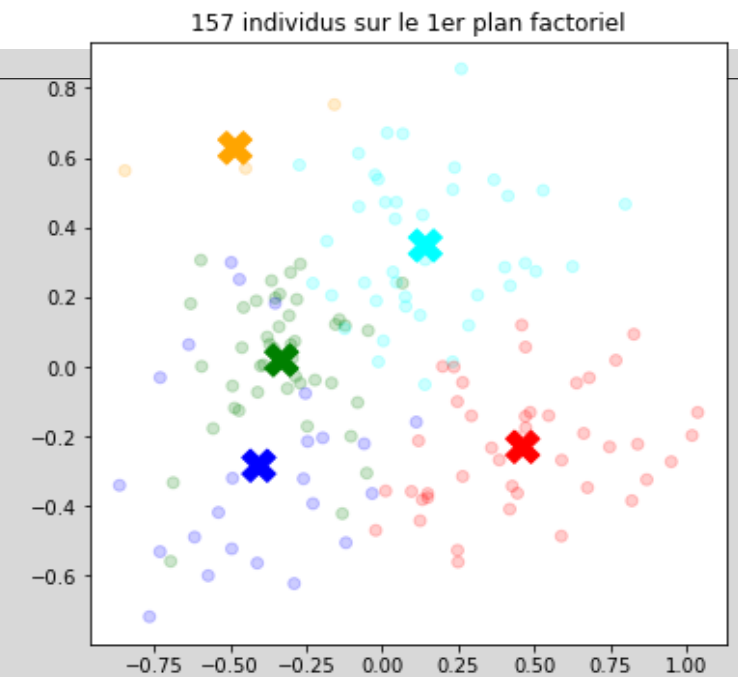
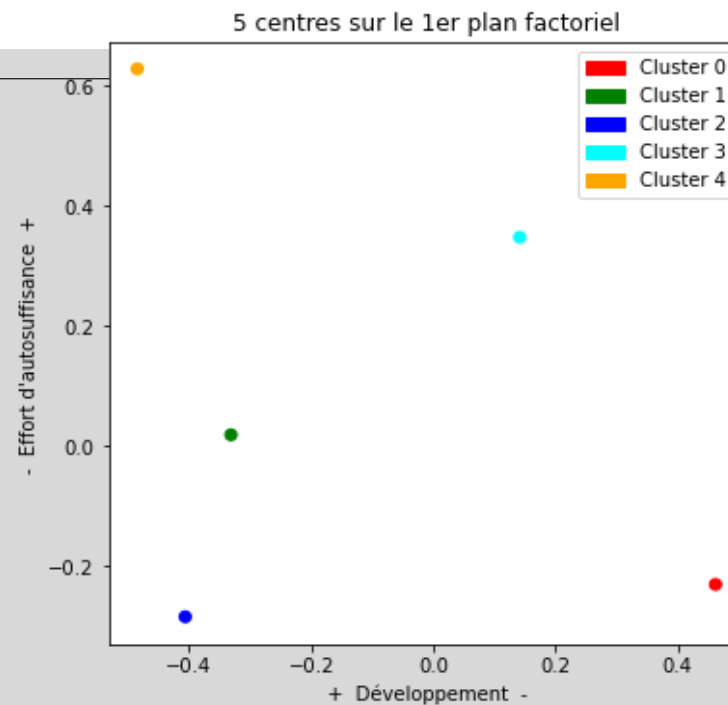
3. K-means

- Spécificité : nécessaire de fixer au préalable le nombre de cluster
- coefficient de silhouette propose des valeurs très similaires (entre 0.17 et 0.195) suivant le nombre de cluster
- Méthode du coude : légère cassure à partir de 5 clusters.



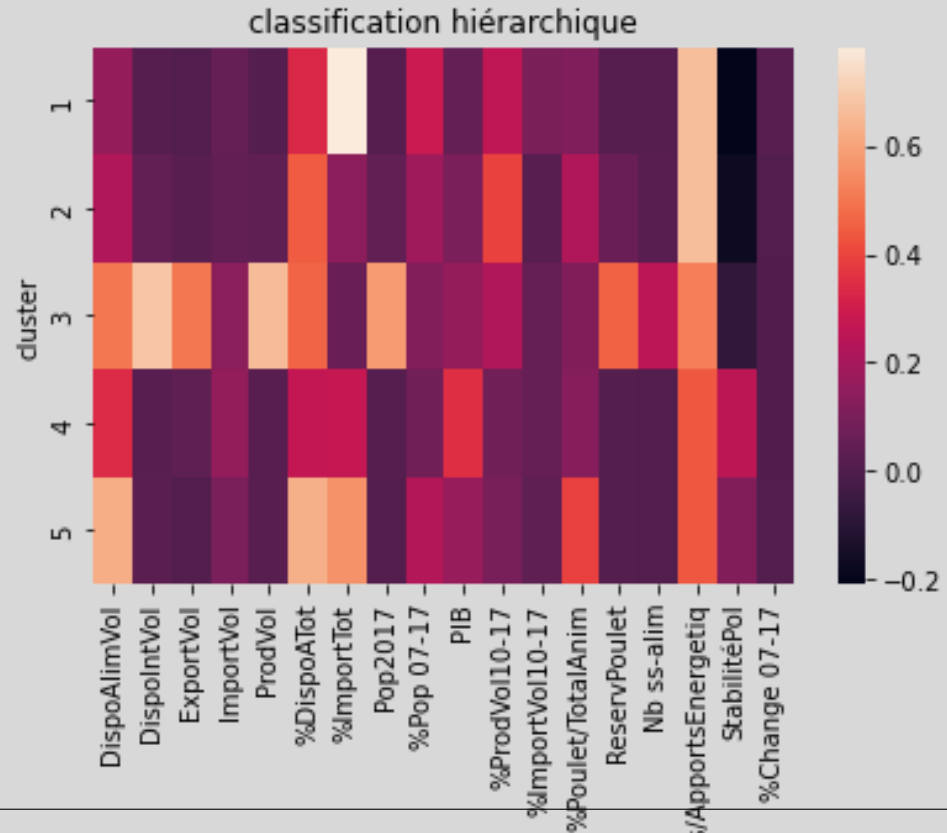
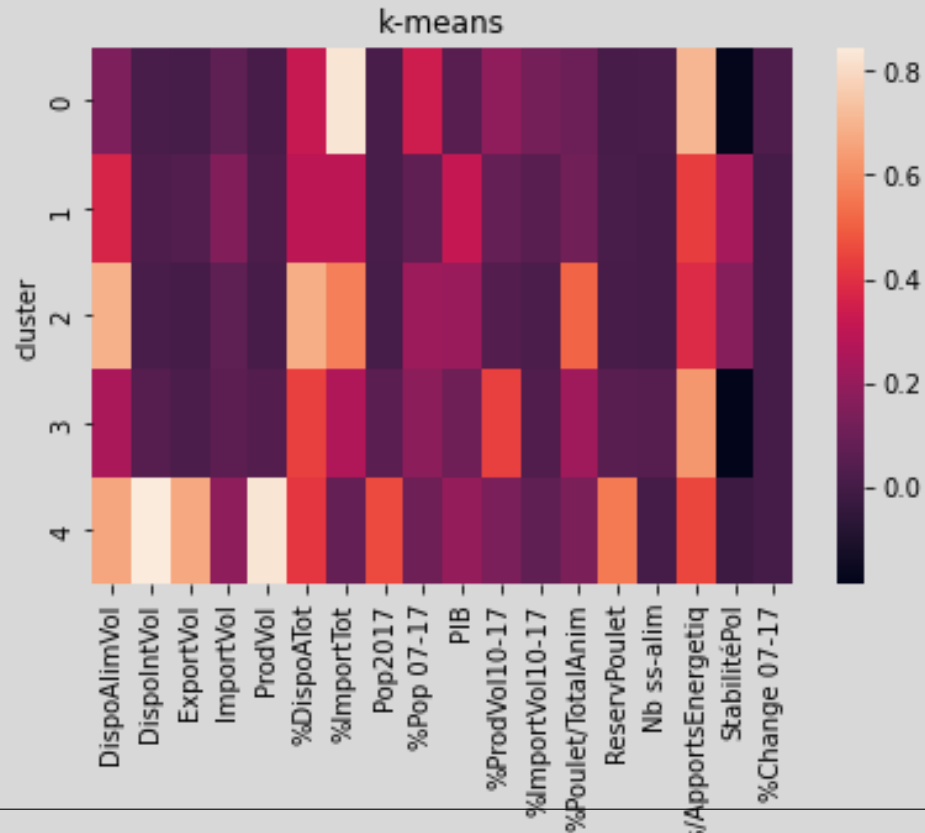
Clustering K-means

Projection des
centroïdes sur les 2
premiers plans
factoriels



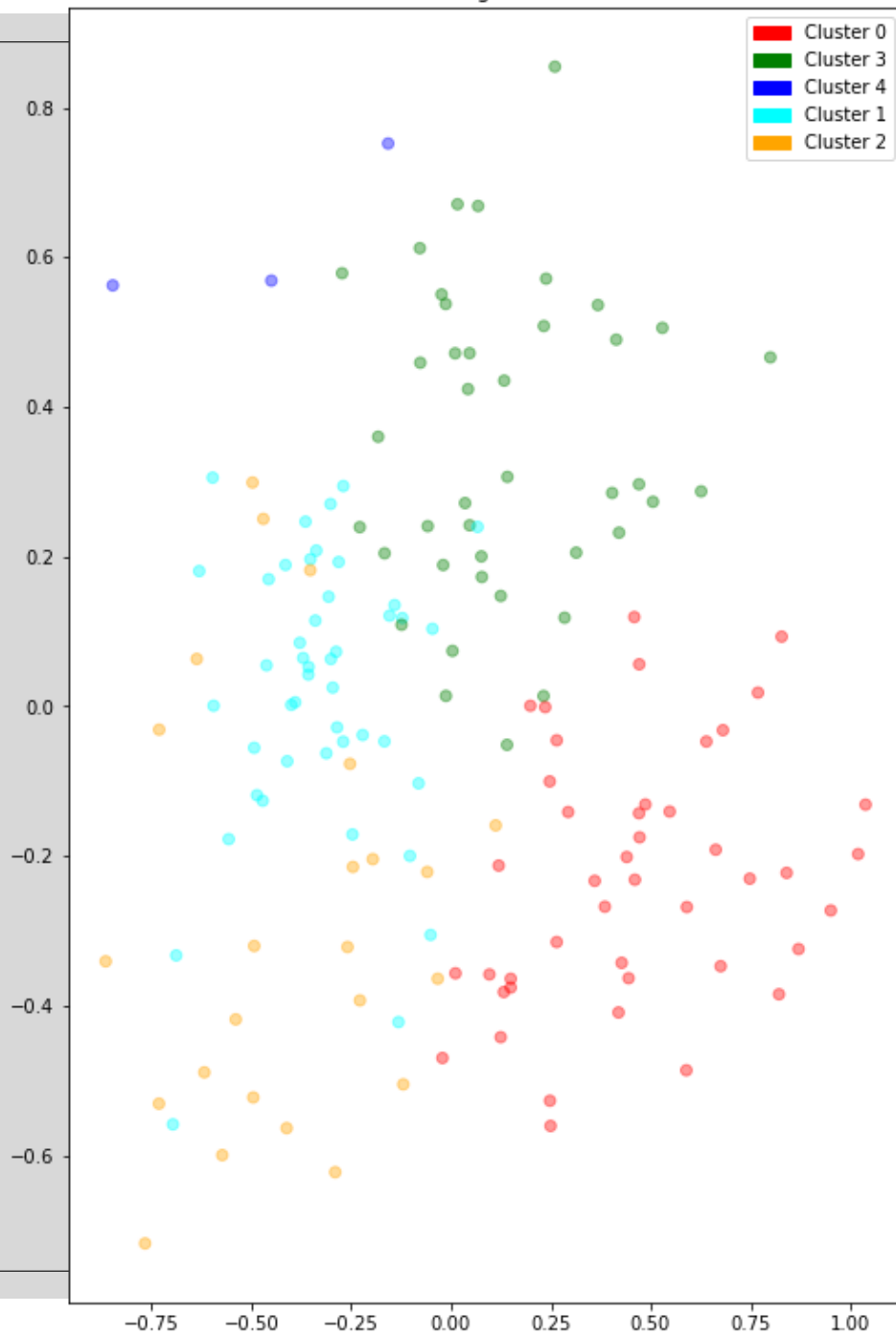
Comparaison des caractéristiques des clusters

- Les caractéristiques des clusters sont similaires quelque soit la méthode choisie
- Quelques changements à la marge :
 - ⇒ cluster K4 – CAH3 : la méthode k-means exclue l'Inde, ce qui a pour effet d'accentuer l'intensité de certaines variables

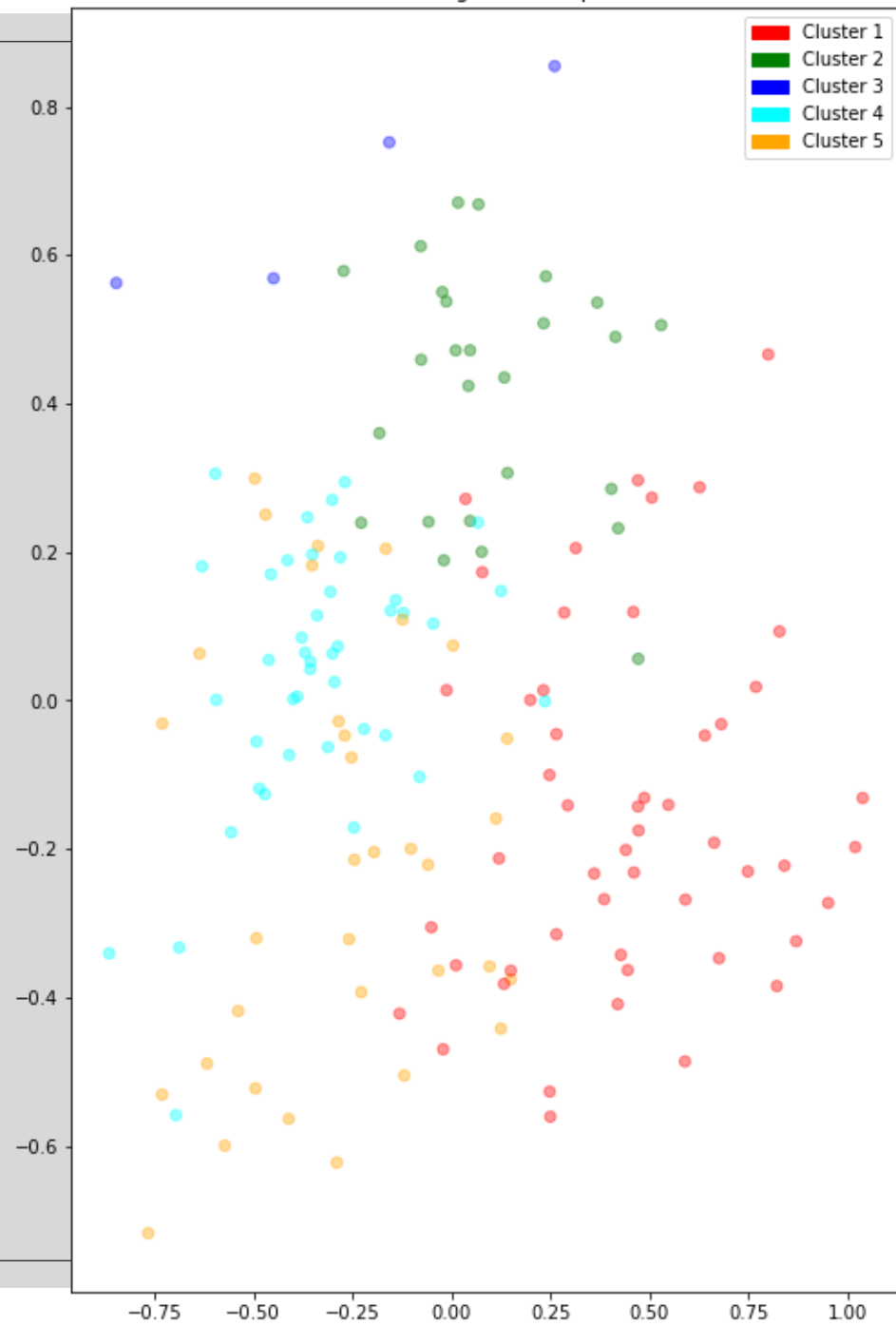


Comparaison

Clustering k-means

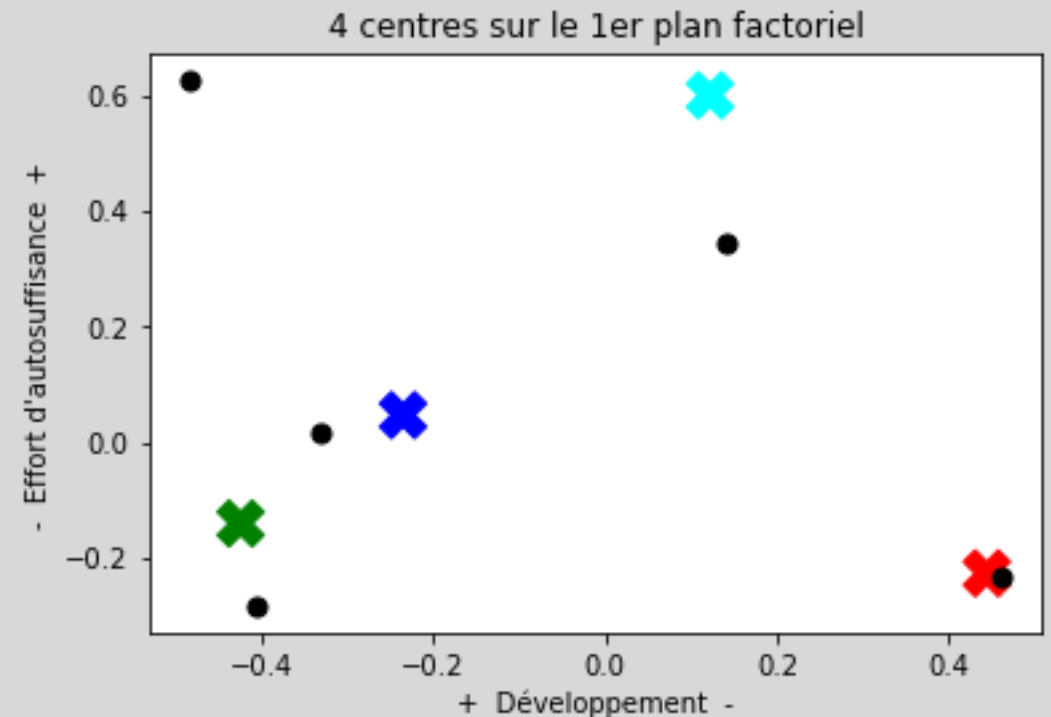


Clustering hiérarchique



Annexe : clustering hors outliers

- Outliers :
 - ⇒ Etats-Unis, Chine, Brésil
 - ⇒ Inde, Zambie, Venezuela
- Synthèse F1 – F4 :
 - ⇒ F1 : + Développement –
 - ⇒ F2 : - Effort d'autosuffisance +
 - ⇒ F3 : - **Spécialisation sur la volaille** +
 - ⇒ F4 : + Dynamique de production de volaille –
- Quelques différences sur le positionnement des centroïdes



Conclusion

- Proposition stratégique pour l'exportation des poulets

- Priorisation des clusters 4 et 5

- ⇒ Les plus stables politiquement et économiquement

- ⇒ autres que cluster 3 (plus grands producteurs mondiaux)

- 2 logiques

- ⇒ cluster 4 : anticipation des besoins futurs (faible niveau actuel de production et de consommation)

- ⇒ cluster 5 : couverture des besoins actuels (haut niveau de consommation et d'importation)

- Cluster 4 :

- Allemagne, Belgique, Pays-Bas :

- ⇒ Part de la dispo. alimentaire < 20%,

- ⇒ Facilité d'échange : Espace Schengen

- Japon, Corée, Hong-Kong :

- ⇒ Demande forte en produits de qualité (prix)

- ⇒ Possible saturation de l'outil de production ?

- Cluster 5 :

- Emirats arabes Unis, Oman, Arabie saoudite, Koweït :

- ⇒ Volumes de production < Volumes importés

- ⇒ Dynamique démographique

- ⇒ Régime alimentaire pour la majorité de la population écarte une autre viande

