



La poule qui chante

P9- Produisez une étude de marché avec Python

SOMMAIRE

- 1- Contexte du projet**
- 2- Préparation des données**
- 3- L'ACP**
- 4- Cercle de corrélation**
- 5- Classification ascendante hiérarchique**
- 6- Clustering par Kmeans**
- 7- Listes des pays retenus**
- 8- Conclusion**

1- Contexte du projet

La poule qui chante, est une entreprise Française spécialisée dans l'agro-alimentaire, plus particulièrement dans la volaille.

Elle ambitionne de développer ses activités à l'international. C'est dans ce cadre qu'il a été demandé de mener une première étude afin de pouvoir cibler les pays potentiels dans lesquels l'entreprise pourrait y lancer ses activités.

Et à la suite de cette étude, une analyse plus fine sera faite plus tard afin de faire un choix final sur le / les pays à retenir.

2- Préparation des données

Initialement 2 fichiers avaient été mis à notre disposition pour effectuer les analyses (Population et Disponibilité Alimentaire).

Mais on avait également la possibilité d'enrichir notre analyse avec des données qu'on estimait utile.

C'est ainsi que les données suivantes ont été ajoutées à notre étude :

- **La distance entre la France et les autres pays**
- **Le coût d'importation de volaille**
- **La Stabilité politique**
- **La taxe sur le commerce international**
- **Le PIB par habitant**
- **Le taux d'utilisation d'internet**

On a également mis en place des champs calculés tel que:

- **Proportion de protéine de volaille**
- **L'évolution de la population entre 2012 et 2017**

2- Préparation des données

Pour chacun des data frames, nous avons effectué les actions suivantes :

- Sélection des colonnes utiles pour nos analyses
- Vérifications des types de données
- Vérification des valeurs manquantes
- Vérification de l'existence de doublon

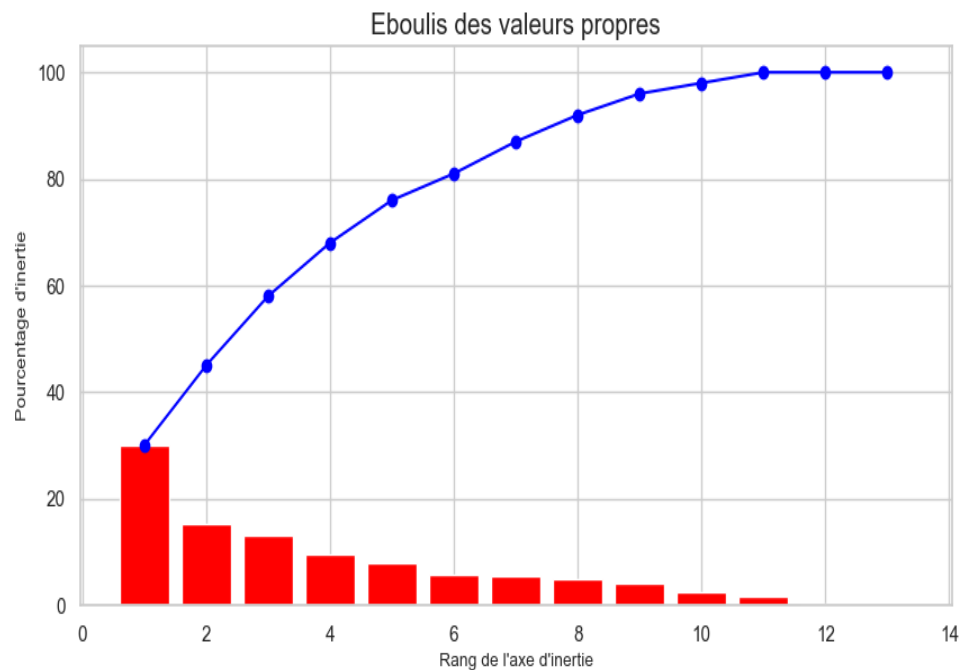
Pour les data frames présentant des valeurs manquantes, les valeurs ont été remplacées par 0 (Ex: Stabilité politique) ou par la valeur moyenne (Ex: Montant importation poulet)

Après le nettoyage des data frames, on a procédé à la fusion de l'ensemble des données sous une donnée unique.

3- L'ACP

L'ACP (Analyse en composantes principales) permet d'étudier des ensembles de données multidimensionnelles avec des variables quantitatives.

Eboulis des valeurs propres afin de déterminer le nombre d'axe à retenir pour notre étude



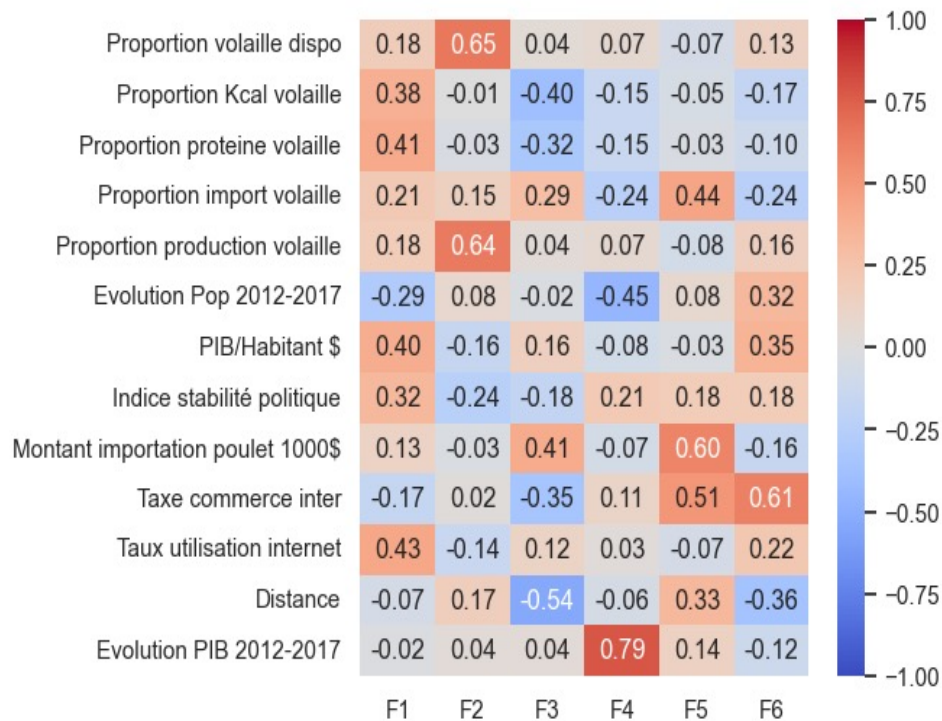
En rouge, nous avons la variance de chaque composante, et en bleu la variance cumulée.

On peut remarquer que 81% de la variance est exprimée par les 6 premières composantes.

Nous allons donc considérer les 6 premières composantes pour notre étude.

3- L'ACP

Heat-map de nos variables et les composantes



Corrélation positive entre l'axe F1 et : La proportion de Kcal de volaille, PIB/Habitant \$, une corrélation négative entre l'évolution de la population entre 2012 et 2017.

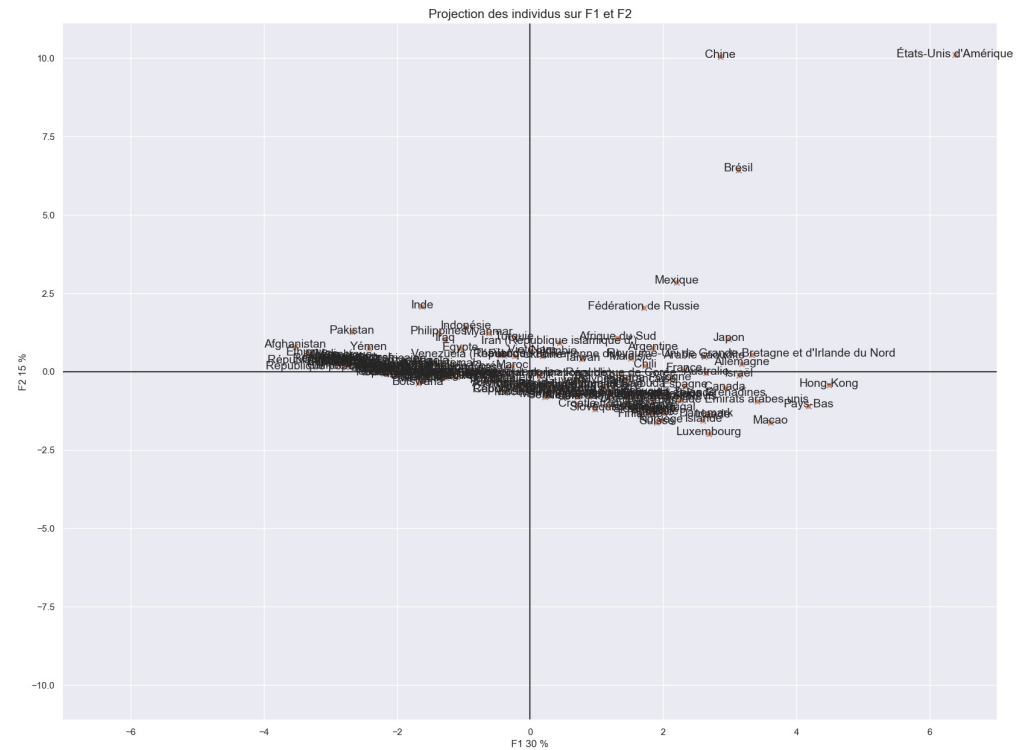
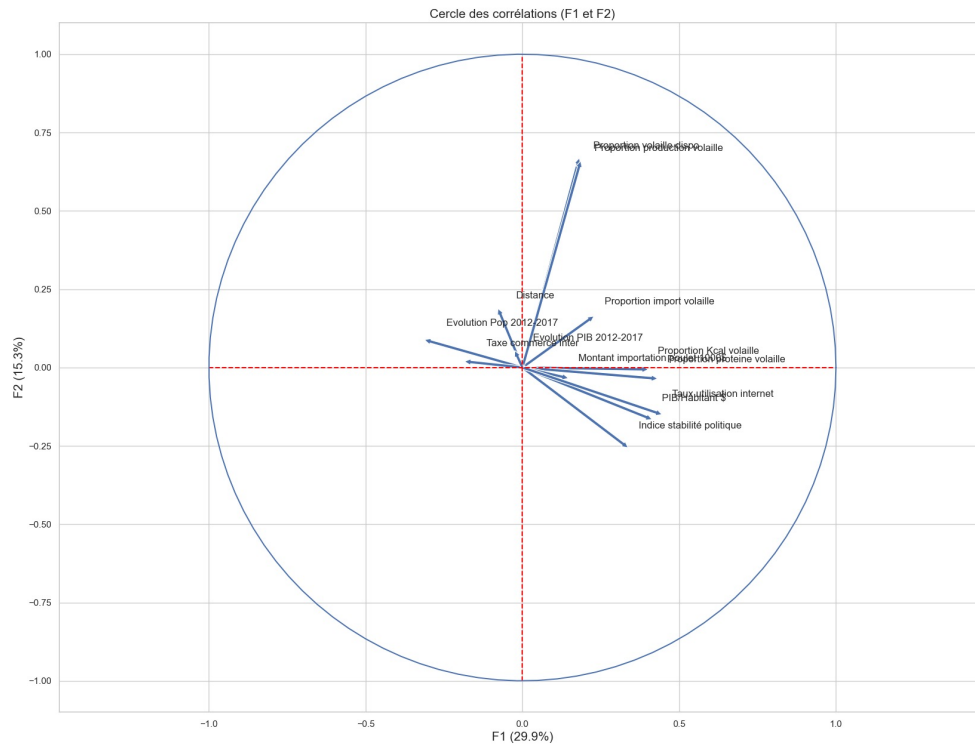
Corrélation positive entre l'axe F2 et : La proportion de volaille dispo. Une corrélation négative avec l'indice de stabilité politique.

Corrélation positive entre l'axe F3 et : La proportion des importations de volaille, une corrélation négative avec le proportion Kcal de volaille.

Corrélation positive entre l'axe F4 et : L'évolution du PIB entre 2012 et 2017 et une corrélation négative avec l'évolution de la population entre 2012 et 2017.

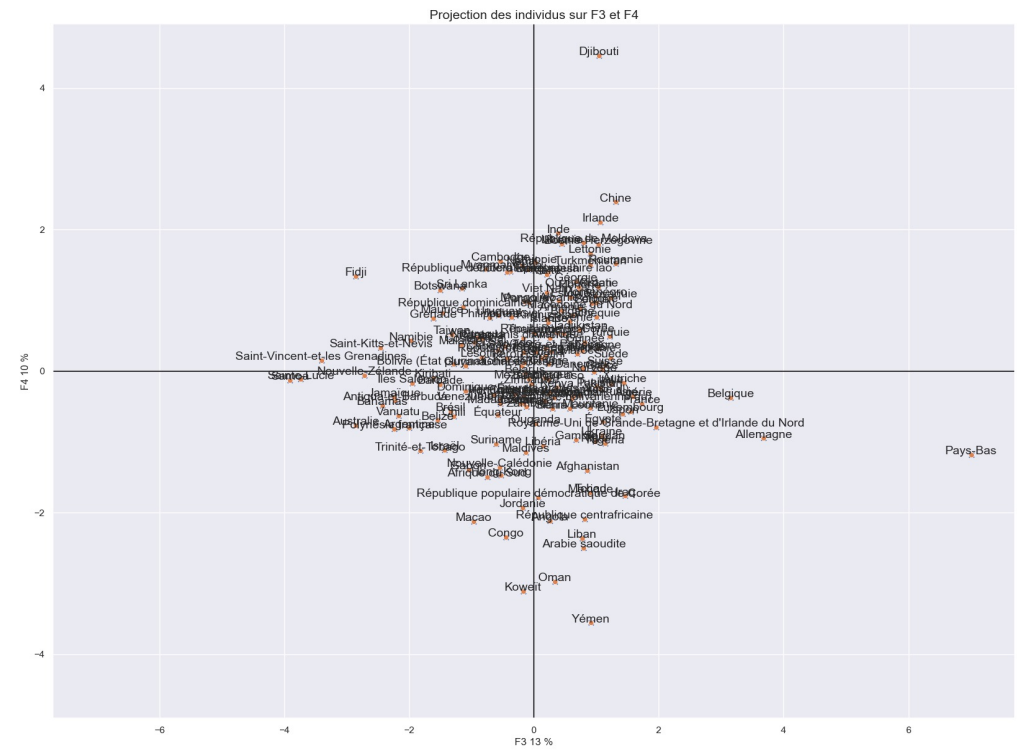
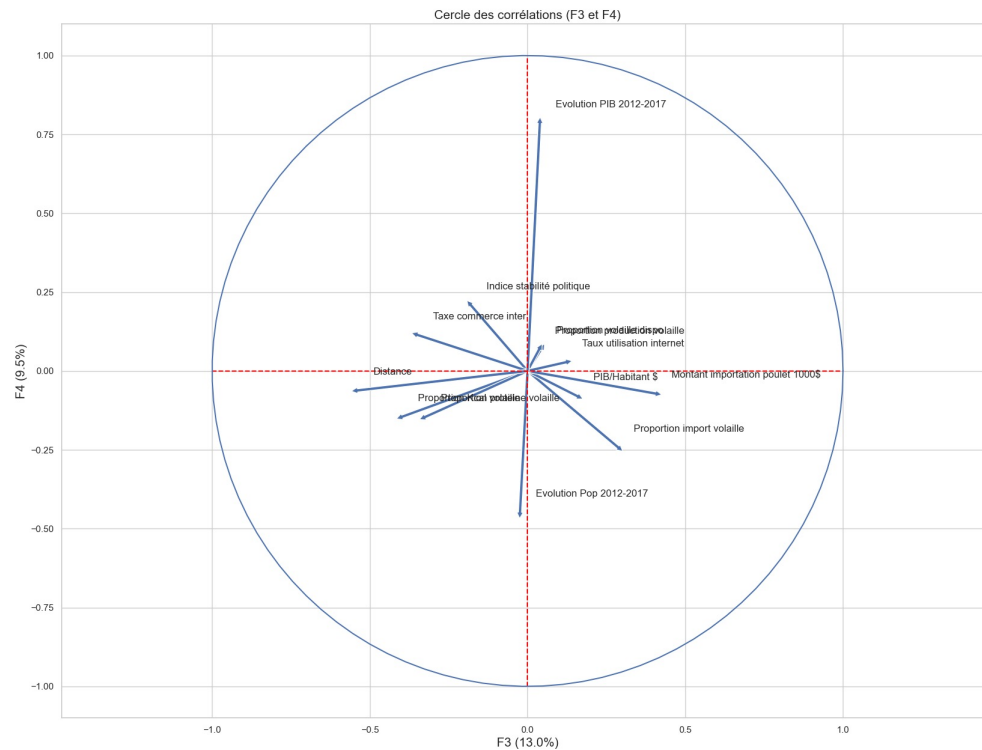
Nous allons chercher à observer les différentes corrélations à travers le cercle de corrélation.

4- Cercle de corrélation



Les pays se trouvant en haut à droite regroupent ceux qui ont une forte proportion de dispo de volaille et de production de volaille.
Les pays en bas à droite regroupent sont ceux qui ont un indice de stabilité politique élevée , un PIB/ habitant important et un taux d'utilisation d'internet

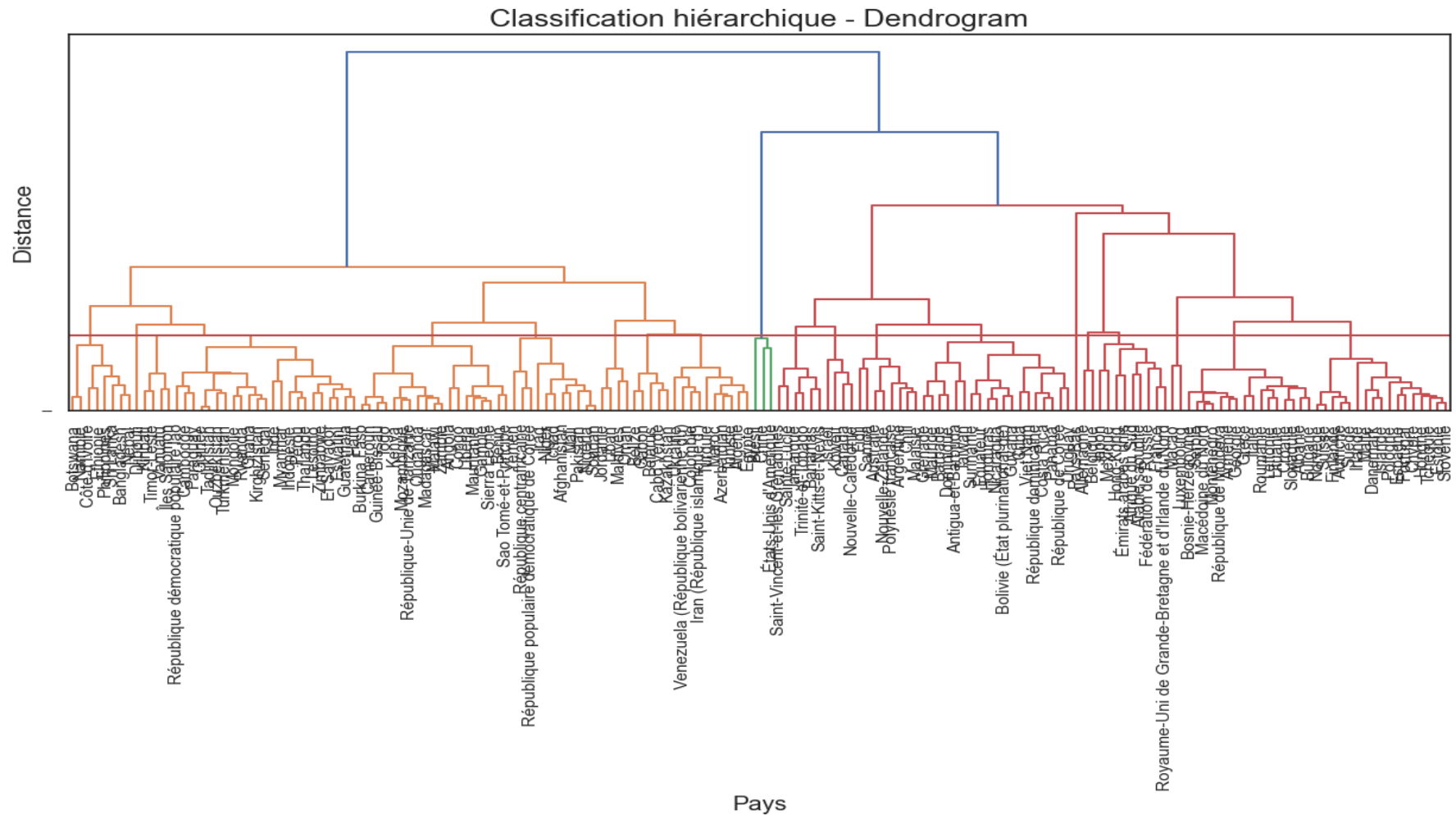
4- Cercle de corrélation



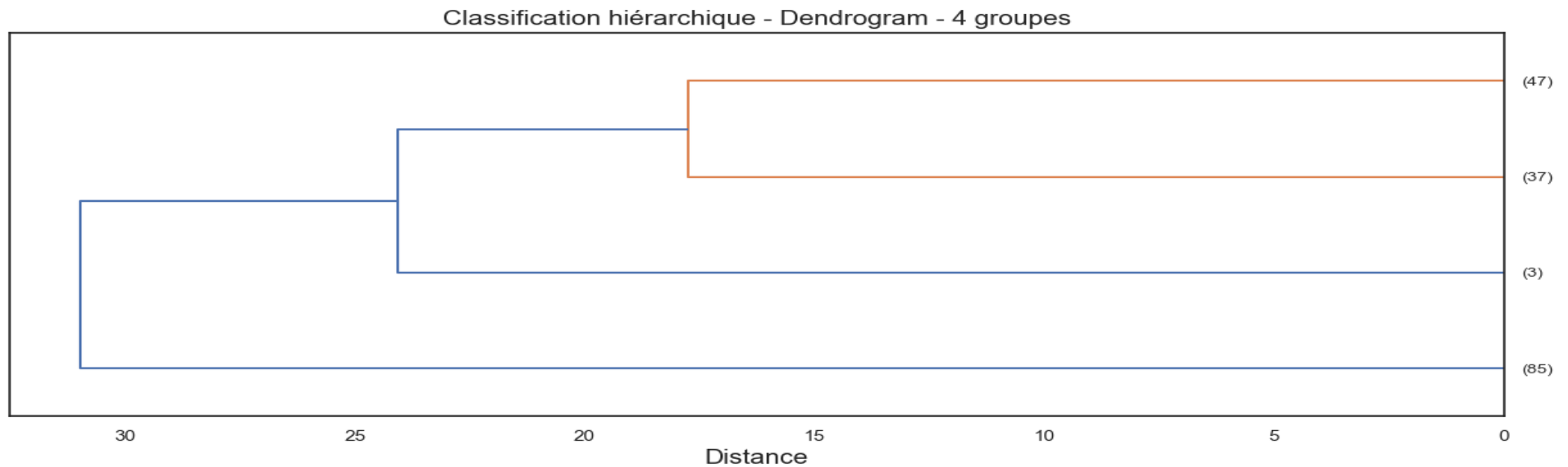
Les pays se trouvant en bas à gauche regroupent ceux qui sont les plus éloignés de la France, qui ont une forte proportion Kcal et de protéine de volaille, ainsi que l'évolution de la population.

Ceux d'en bas à droite regroupent ceux ayant une grande proportion d'importation de volaille, et qui dépensent le plus pour l'importation de volaille.

5- Classification ascendante hiérarchique



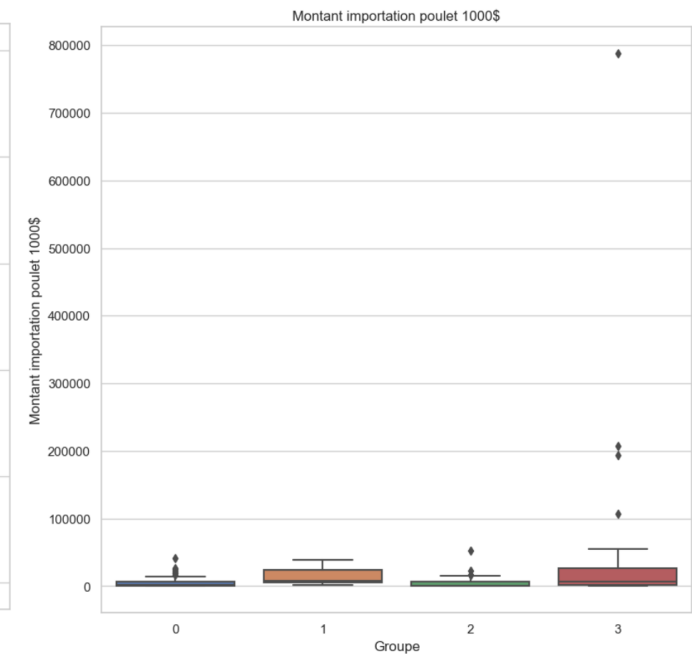
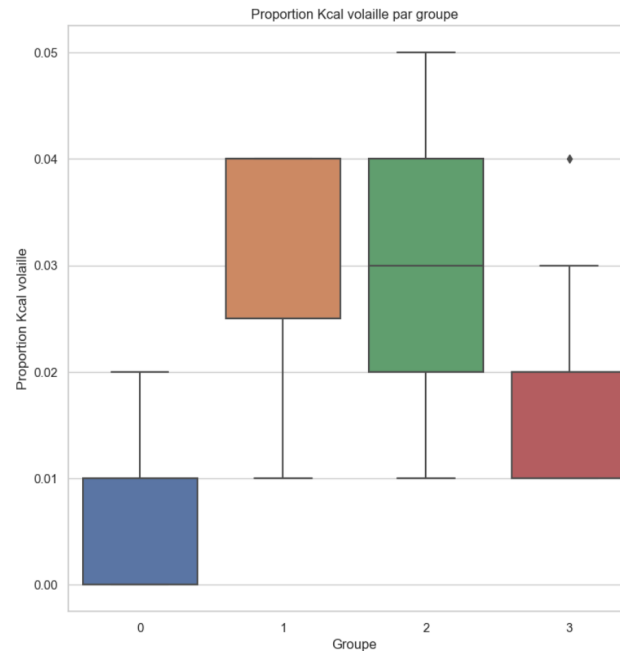
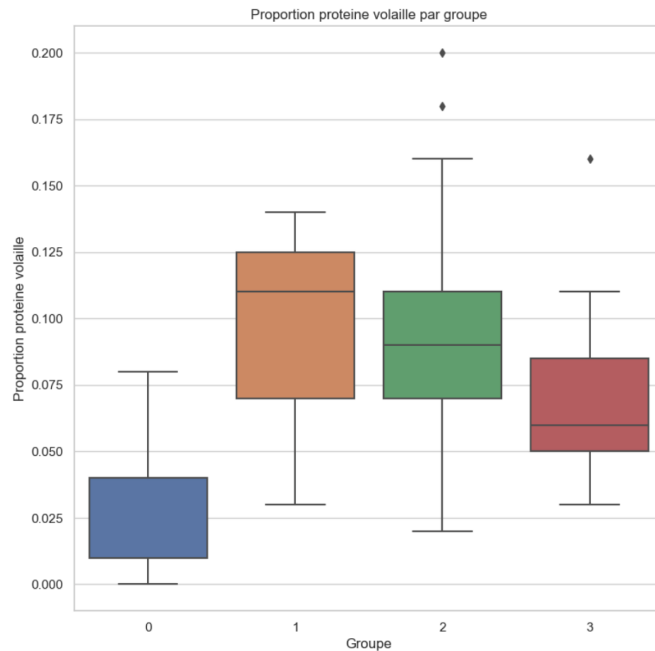
5- Classification ascendante hiérarchique



Après l'essai de plusieurs combinaisons, nous avons opté pour un dendrogramme à 4 groupes.

5- Classification ascendante hiérarchique

Moyenne des groupes / Centroïde

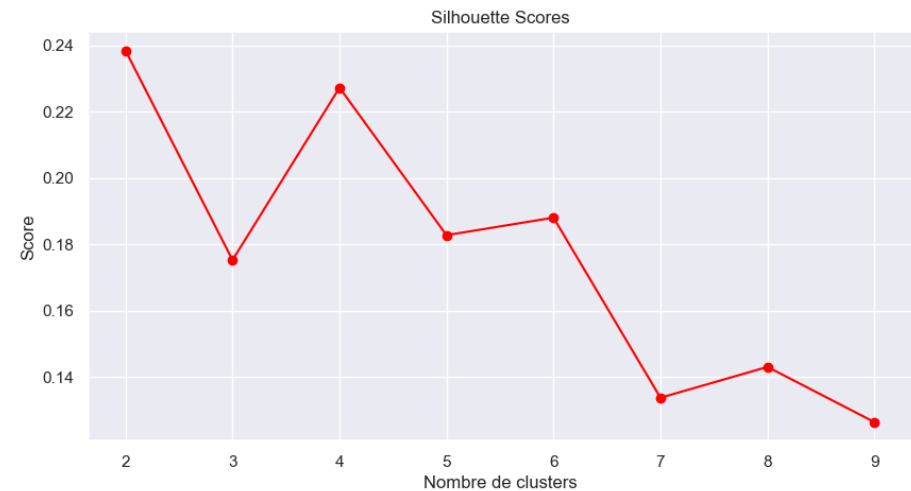
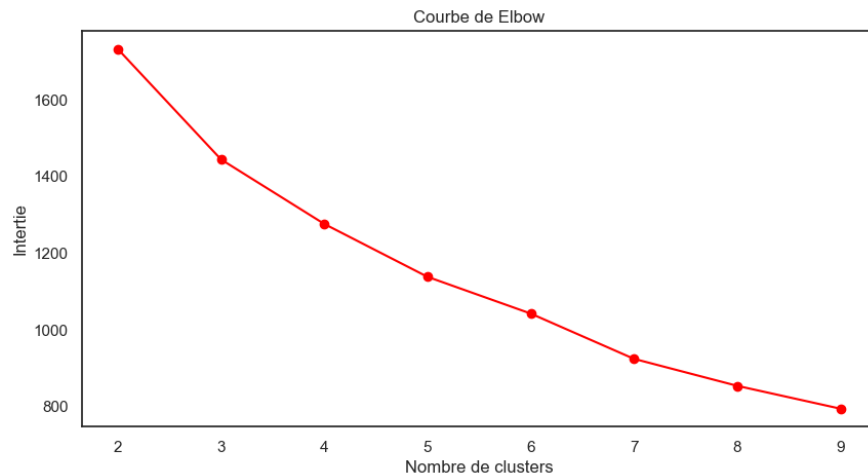


Si on se base sur la proportion de protéine et de proportion de Kcal de volaille, le groupe à retenir est le groupe 2 mais également le groupe 1.

Mais si on se base sur l'importance des dépenses et de la quantité d'importation de volaille, c'est le groupe 3 qui sera retenu.

Il ressort du l'union des groupes 1, 2 et 3, 87 pays que l'on pourrait cibler pour notre entreprise sur la base de la proportion de protéine de volaille et aussi des montants des dépenses.

6- Clustering par Kmeans



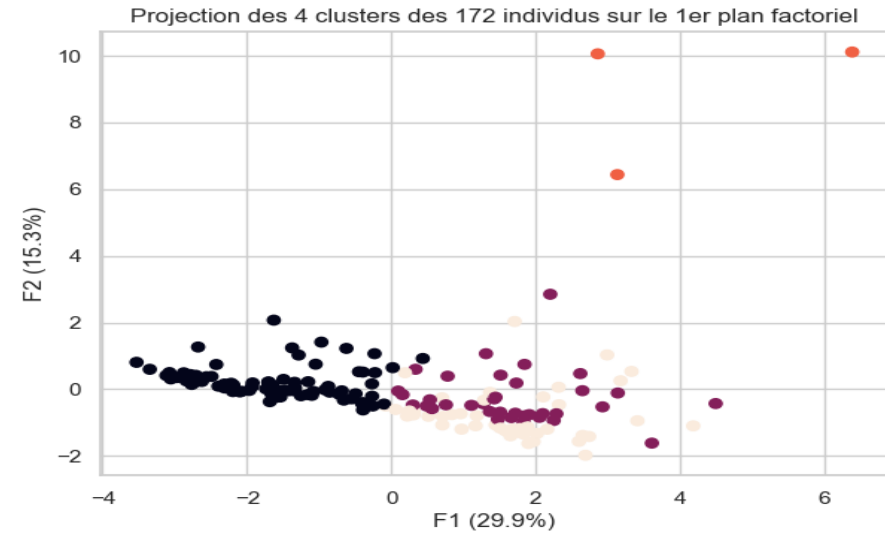
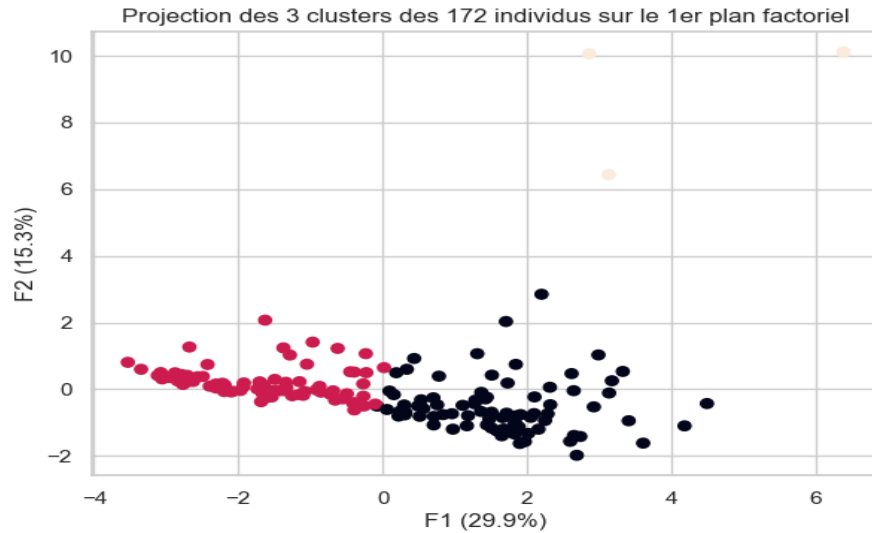
Afin de déterminer le nombre de cluster optimal, nous avons utiliser la Courbe de Elbow ainsi que le Silhouette Score.

Si on se fie à la courbe de Elbow, on va retenir $K = 3$ ou $K = 4$.

Le coefficient de silhouettes avec un $K = 4$ devra également être considéré dans notre analyse. On va donc effectuer un clustering avec $K = 3$ et ensuite $K = 4$ et observer les résultats obtenus.

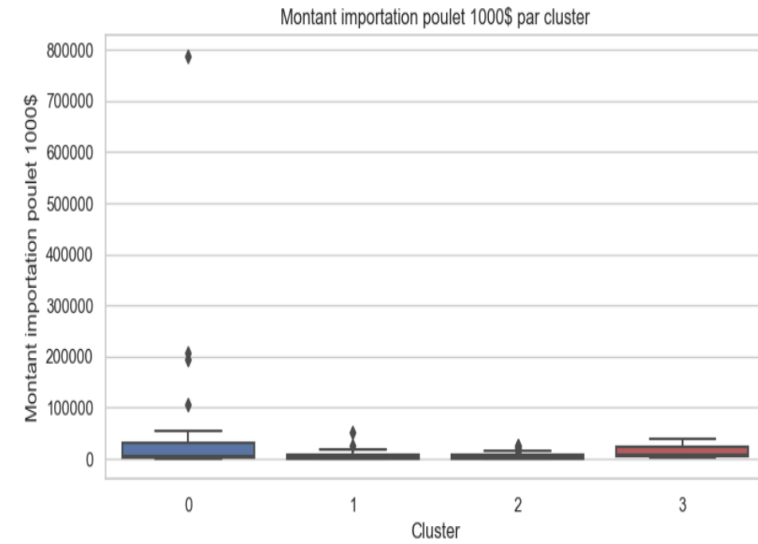
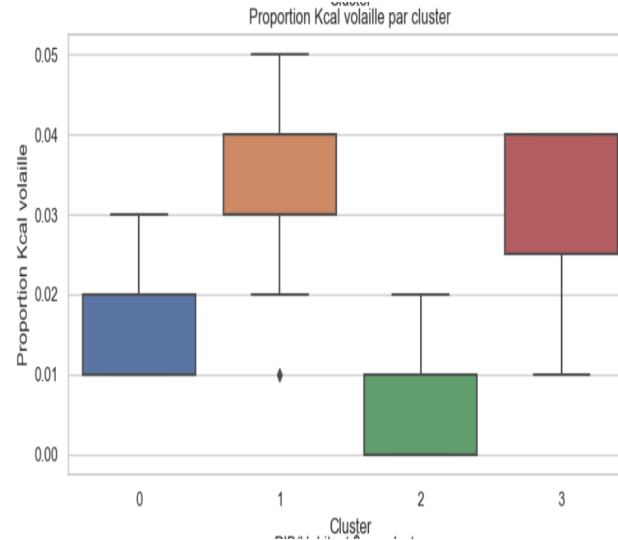
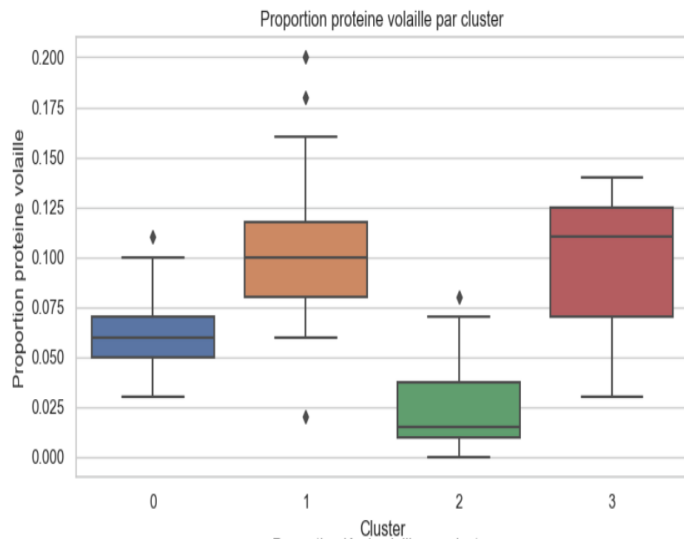
Ainsi avec $K = 4$, cela nous permettra de comparer les résultats avec le découpage effectué avec le dendrogramme.

6- Clustering par Kmeans



On peut observer qu'il est bien possible de faire la projection en 3 clusters, les 3 clusters sont bien distincts. Mais avec 3 clusters, on se retrouve avec un cluster qui a un seul pays. Mais afin d'avoir une analyse un peu plus poussée et éviter d'avoir un cluster contenant qu'un seul pays, nous allons donc continuer notre analyse avec 4 clusters.

6- Clustering par Kmeans



En se basant sur de la proportion de protéine et de proportion de Kcal de volaille, le groupe à retenir est le cluster 1 et aussi 3. on pourrait également ajouter le cluster 0 si on se basant sur les dépenses en importation de volaille. L'union de ces 3 clusters nous donne une liste de 86 pays.

7- Listes des pays retenus

Pays à cibler selon la proportion de protéine de volaille

◆	Zone ◆
42	Israël
71	Saint-Vincent-et-les Grenadines
73	Samoa
18	Hong-Kong
70	Saint-Kitts-et-Nevis
72	Sainte-Lucie
86	États-Unis d'Amérique
3	Antigua-et-Barbuda
7	Australie
81	Trinité-et-Tobago

Pays à cibler selon la Proportion Kcal volaille

◆	Zone ◆
72	Sainte-Lucie
42	Israël
3	Antigua-et-Barbuda
71	Saint-Vincent-et-les Grenadines
86	États-Unis d'Amérique
44	Jamaïque
18	Hong-Kong
34	Grenade
70	Saint-Kitts-et-Nevis
14	Brésil

Pays à cibler selon la Proportion import volaille

◆	Zone ◆
54	Mexique
45	Japon
2	Allemagne
18	Hong-Kong
4	Arabie saoudite
66	Royaume-Uni de Grande-Bretagne et d'Irlande du...
61	Pays-Bas
32	France
21	Chine
0	Afrique du Sud
84	Émirats arabes unis

Pays à cibler selon le PIB/Habitant \$

◆	Zone ◆
19	Macao
49	Luxembourg
40	Irlande
77	Suisse
84	Émirats arabes unis
57	Norvège
86	États-Unis d'Amérique
18	Hong-Kong
41	Islande
26	Danemark

7- Listes des pays retenus

Pays à cibler selon le Montant importation poulet 1000\$

◆	Zone ◆
61	Pays-Bas
11	Belgique
2	Allemagne
62	Pologne
39	Hongrie
16	Canada
28	Espagne
65	Roumanie
66	Royaume-Uni de Grande-Bretagne et d'Irlande du...
21	Chine

Pays à cibler selon la Distance entre la France et les autres pays (Pays les plus proche)

◆	Zone ◆
32	France
11	Belgique
49	Luxembourg
66	Royaume-Uni de Grande-Bretagne et d'Irlande du...
61	Pays-Bas
77	Suisse
40	Irlande
2	Allemagne
80	Tchéquie
76	Slovénie
26	Danemark

Pays à cibler selon l'Indice stabilité politique

◆	Zone ◆
59	Nouvelle-Zélande
19	Macao
41	Islande
49	Luxembourg
77	Suisse
52	Malte
27	Dominique
73	Samoa
57	Norvège
64	Portugal

8- Conclusion

Au terme de nos différentes explorations, analyses et visualisations, Nous avons retenu 4 clusters avec le Dendrogramme et la méthode du Kmeans.

Il ressort du rapprochement de ces deux méthodes une première liste de 82 pays.

Afin de réduire cette liste, il faut avec les spécialistes du métier se mettre d'accord sur les critères les plus importants et pertinents à prendre en compte afin d'affiner les listes obtenues.

Ainsi, après le choix du/des critères afin d'avoir une liste réduite des pays, il va ensuite falloir se pencher sur l'aspect culturel des pays cibles, afin de mettre en place les stratégies adéquates pour faciliter l'accès aux marchés des pays cible et également former nos collaborateurs en cas de barrière de la langue ou les contraintes juridiques spécifiques a chaque pays.