

KAMEL BADAR
DATA ANALYST
22/04/2023



La poule qui chante

Etude de marché

Recommandations sur les opportunités
d'installation à l'étranger



Sommaire

01

Contexte du
projet

02

Démarche

03

Résultats

04

Recommanda
tions

Contexte



Production d'une étude de marché

Ce projet vise à produire une étude de marché afin d'évaluer les opportunités de développement à l'international pour notre entreprise. Il s'agit, à partir d'une analyse macro-économique, d'étudier les opportunités en ce qui concerne le marché de la volaille à l'international.

Contexte

P	E	S	T	E	L
Politique	Economie	Social	Technologie	Environnement	Legal
Stabilité gouvernement Corruption Fiscalité Liberté de la presse Réglementation Tarifs spéciaux Force syndicales Règles concurrence Participation des électeurs Manifestations Niveau des subventions Relations bilatérales Importations / exportations Contrôle du commerce Lobbying Budgets publics Facteurs économiques	Taux de croissance Taux d'intérêt Taux d'inflation Taux de change Crédit Revenu disponible Niveau dépense Déficits budgétaires Évolution du PIB Chômage Marché boursier Fluctuations des prix Facteurs sociaux	Taille population Taux de natalité / mortalité Mariages / divorces Immigration / émigration Espérance de vie Répartition richesse Classes sociales Revenu / habitant Modes de vie Santé Revenu disponible Attitude / travail Habitudes d'achat Education Minorités Epargne Technologies	Incitations technologiques Automatisation Activité de R&D Changement technologique Accès nouvelles technologies Innovation Conscience technologique Infrastructure Internet Infrastructure communication Cycle de vie de la technologie	Météo Climat Politiques environnementales Changement climatique Pressions ONG Catastrophes naturelles Pollution Recyclage Produits verts Energies renouvelables	Discrimination Lois antitrust Droit du travail Protection consommateurs Droit d'auteur & brevets Lois santé & sécurité Lois sur l'éducation Protection des données

Démarche

Choix des variables PESTEL / Sources de données

- **Politique** : indice de stabilité politique (*source: FAO*)
- **Economie** : PIB par habitant (*source: FAO*)
- **Social** : consommation de volailles , population (*source: FAO*)
- **Technologie** : population ayant accès à internet (*source: World Bank*)
- **Environnement** : distance depuis la France (*source: <https://github.com/mlschneid/country-distance/blob/master/output/distance-matrix.csv>)*)
- **Légal** : pourcentage de revenu issu des taxes sur le commerce (*source: World Bank*)

Les données seront importées sur VSCode et le langage utilisé sera Python

Démarche : Exploration

Etape 1 : Import et préparation des données

	Total dispo alim	Poulet dispo alim	% Volaille / Dispo alim (kg/pers/an)
Zone			
Afghanistan	357.31	1.53	0.43
Afrique du Sud	556.42	35.69	6.41
Albanie	1260.49	16.36	1.30
Algérie	798.71	6.38	0.80
Allemagne	943.24	19.47	2.06
...
Émirats arabes unis	627.67	43.47	6.93
Équateur	563.62	19.31	3.43
États-Unis d'Amérique	990.38	55.68	5.62
Éthiopie	361.77	0.13	0.04
Îles Salomon	659.79	4.45	0.67

Préparer les données de sorte à aboutir à un pivot table par champs du PESTEL qui contient les différentes variables sélectionnées

Démarche : Exploration

Etape 2 : Jointures et jeu de données

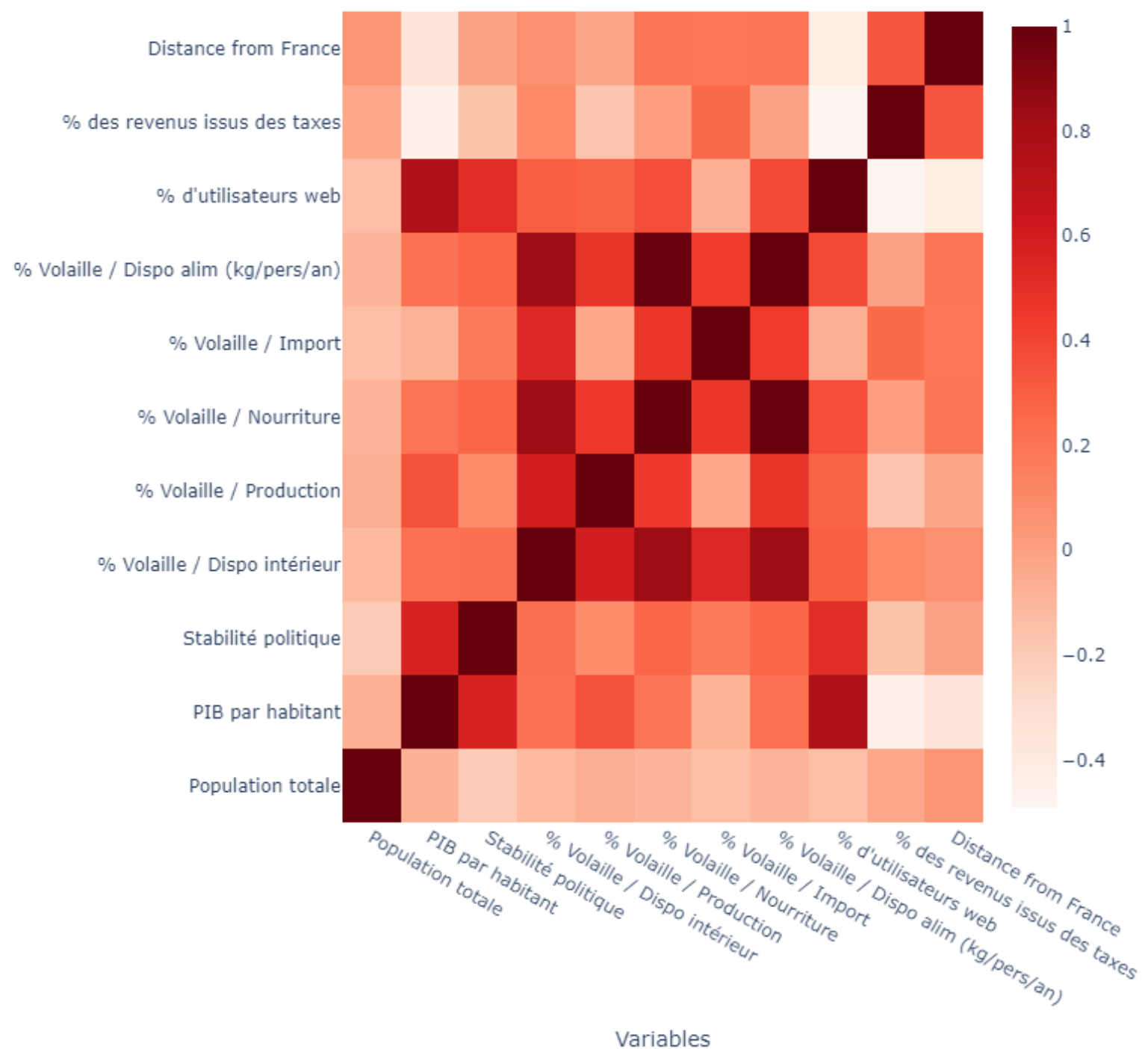
	Zone	ISO3	Population totale	PIB par habitant	Stabilité politique	% Volaille / Dispo intérieur	% Volaille / Production	% Volaille / Nourriture	% Volaille / Import	% Volaille / Dispo alim (kg/pers/an)	% d'utilisateurs web	% des revenus issus des taxes	Distance from France
0	Afghanistan	AFG	36296113	2058.4	-2.80	0.38	0.24	0.42	0.64	0.43	13.50	7.42	5594.0
1	Afrique du Sud	ZAF	57009756	12701.3	-0.28	3.17	2.44	6.41	5.70	6.41	56.17	3.17	9307.0
2	Albanie	ALB	2884169	12771.0	0.38	0.96	0.31	1.29	4.30	1.30	62.40	1.59	1604.0
3	Algérie	DZA	41389189	11737.4	-0.92	0.61	1.03	0.80	0.01	0.80	47.69	NaN	1345.0
4	Allemagne	DEU	82658409	52952.9	0.59	0.99	0.90	2.06	1.27	2.06	84.39	0.00	881.0
...
169	Émirats arabes unis	ARE	9487203	67183.6	0.62	3.49	2.68	6.92	2.89	6.93	94.82	0.08	5252.0
170	Équateur	ECU	16785361	11617.9	-0.07	1.43	1.10	3.42	0.00	3.43	55.80	NaN	9626.0
171	États-Unis	USA	325084756	60109.7	0.34	2.15	2.25	5.62	0.15	5.62	87.27	1.01	5853.0
172	Éthiopie	ETH	106399924	2021.6	-1.68	0.03	0.03	0.04	0.03	0.04	18.62	18.74	5566.0
173	Îles Salomon	SLB	636039	2663.9	0.20	0.30	0.00	0.72	4.84	0.67	11.92	19.72	15164.0

Jointures successives de chaque tables pour aboutir au jeu de données qui contient toutes les variables sélectionnées pour le PESTEL

Démarche : Exploration

Etape 3 : Analyse des variables

Matrice de corrélation



La matrice de corrélation permet ici de mettre en évidence les variables fortement corrélées pour éviter la redondance dans notre dataset

Nous excluons donc de l'analyse les variables suivantes :

"Volaille / Dispo intérieur", "% Volaille / Production",

"% Volaille / Dispo alim (kg/pers/an)"

Démarche : Modélisation

Etape 1 : Normalisation des données

Il est nécessaire de réaliser un scaling sur ce jeu de données car les variables sont sur des échelles différentes. Nos analyses précédentes ont révélé la présence d'outliers donc nous choisissons la méthode de scaling MaxAbsScaler

Formule :

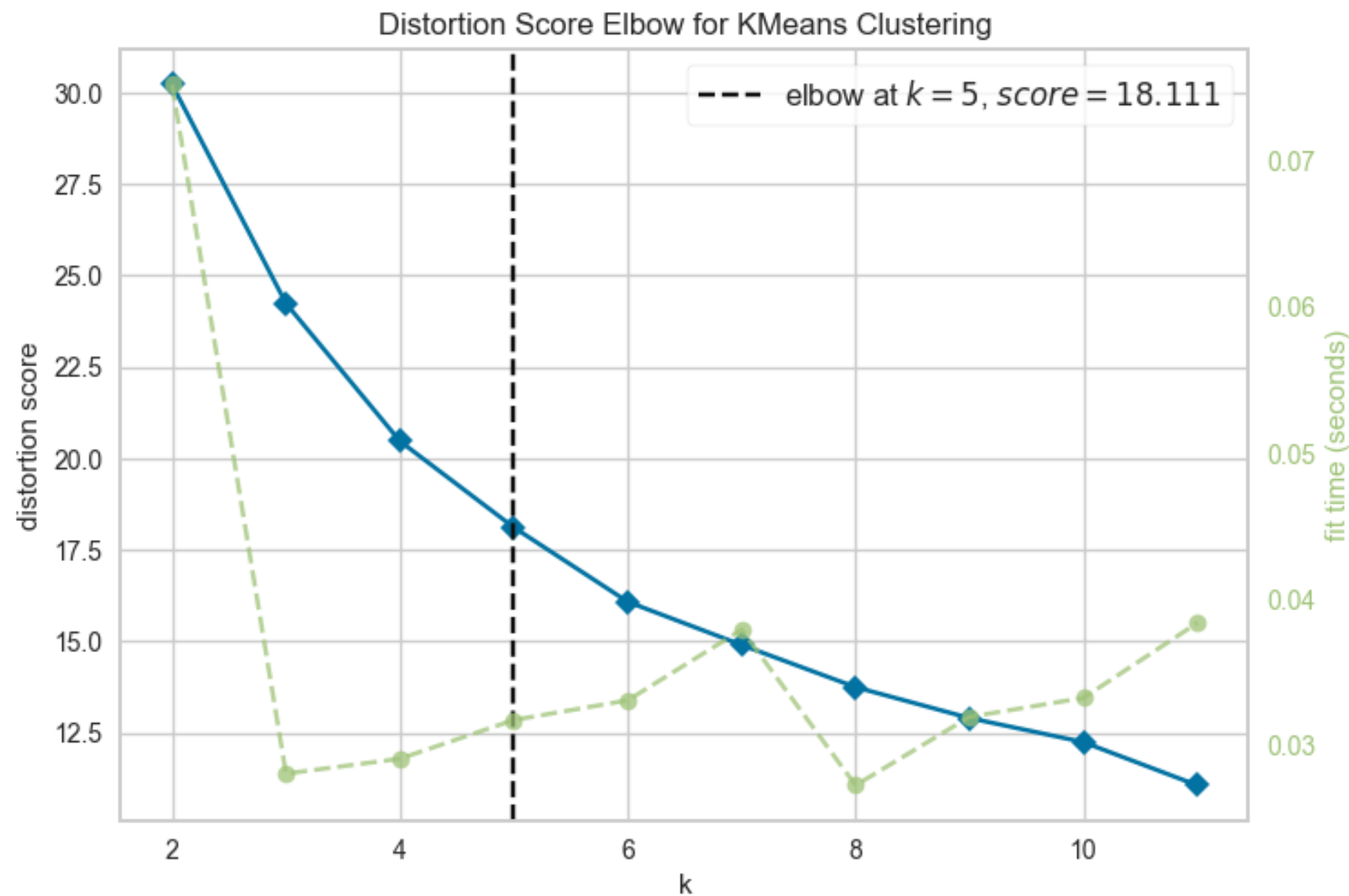
$$\mathbf{x_scaled} = \mathbf{x} / \max(\mathbf{abs(x)})$$

	Population totale	PIB par habitant	Stabilité politique	% Volaille / Nourriture	% Volaille / Import	% d'utilisateurs web	% des revenus issus des taxes	Distance from France
Zone								
Afghanistan	0.027113	0.016313	-1.000000	0.032209	0.026348	0.137391	0.228448	0.301726
Afrique du Sud	0.042587	0.100657	-0.100000	0.491564	0.234664	0.571647	0.097599	0.501996
Albanie	0.002154	0.101209	0.135714	0.098926	0.177028	0.635050	0.048953	0.086516
Allemagne	0.061746	0.419648	0.210714	0.157975	0.052285	0.858844	0.000000	0.047519
Angola	0.022273	0.057938	-0.117857	0.141871	0.274599	0.264604	0.111761	0.349461

Voici les données finales permettant d'effectuer le clustering

Démarche : Modélisation

Etape 2 : Clusters optimaux kmeans



For 2 clusters, the silhouette score is 0.3133
For 3 clusters, the silhouette score is 0.3103
For 4 clusters, the silhouette score is 0.2757
For 5 clusters, the silhouette score is 0.2898
For 6 clusters, the silhouette score is 0.2723
For 7 clusters, the silhouette score is 0.2829
For 8 clusters, the silhouette score is 0.2349
For 9 clusters, the silhouette score is 0.2646
For 10 clusters, the silhouette score is 0.2264
For 11 clusters, the silhouette score is 0.2738
For 12 clusters, the silhouette score is 0.2932

L'analyse du nombre de clusters optimaux aboutit au choix de 5 clusters

Démarche : Modélisation

Etape 3 : Analyse des clusters kmeans

Visualisation des centroïdes en heatmap



On compare les centroïdes des clusters avec les valeurs médianes pour chaque variable

Démarche : Modélisation

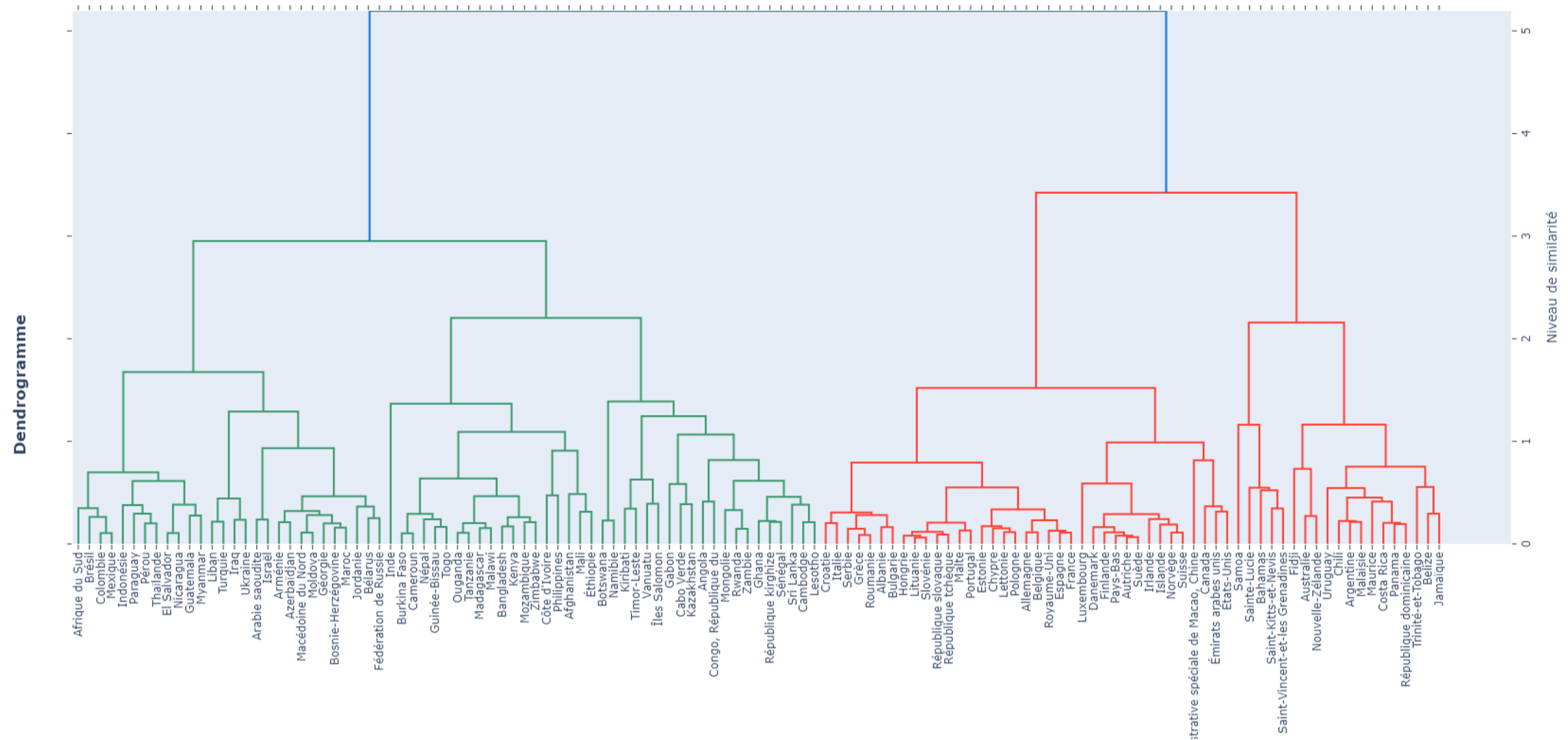
Etape 3 : Analyse des clusters kmeans

- **Cluster 0 (23 pays)** : Valeurs très proches des médianes pour la plupart des centroïdes sauf pour la stabilité politique. Ces pays représentent un marché intéressant notamment pour les données relatives à la consommation de volaille mais ils sont très instables politiquement.
- **Cluster 1 (37 pays)** : Pays possédant la plus forte population mais avec un PIB par habitant très faible. Ils sont également instables politiquement et ne sont pas des grands consommateurs de volaille. Rien d'intéressant non plus pour le web et les taxes, et ce sont des pays plutôt éloignés de la France.
- **Cluster 2 (20 pays)** : Valeurs proches des médianes mais pays qui affichent la plus grande distance par rapport à la France.
- **Cluster 3 (34 pays)** : Pays ayant une population plutôt élevée avec la plus haute stabilité politique une proximité avec la médiane pour les éléments relatifs à la consommation de produits volaillers. On observe un accès au Web très développé, un faible niveau de taxe et une proximité avec la France.
- **Cluster 4 (9 pays)** : Pays peu peuplés avec un PIB/hab et une stabilité médians. On observe que ce cluster regroupe les plus gros consommateurs de poulet mais ils sont éloignés de la France et peu ouvert au commerce avec des taxes élevées.

Le cluster 3 est le plus intéressant ici et sera retenu pour la suite. Bien que présentant certains points positifs, les autres seront écartés.

Démarche : Modélisation

Etape 2 : Clusters optimaux CAH

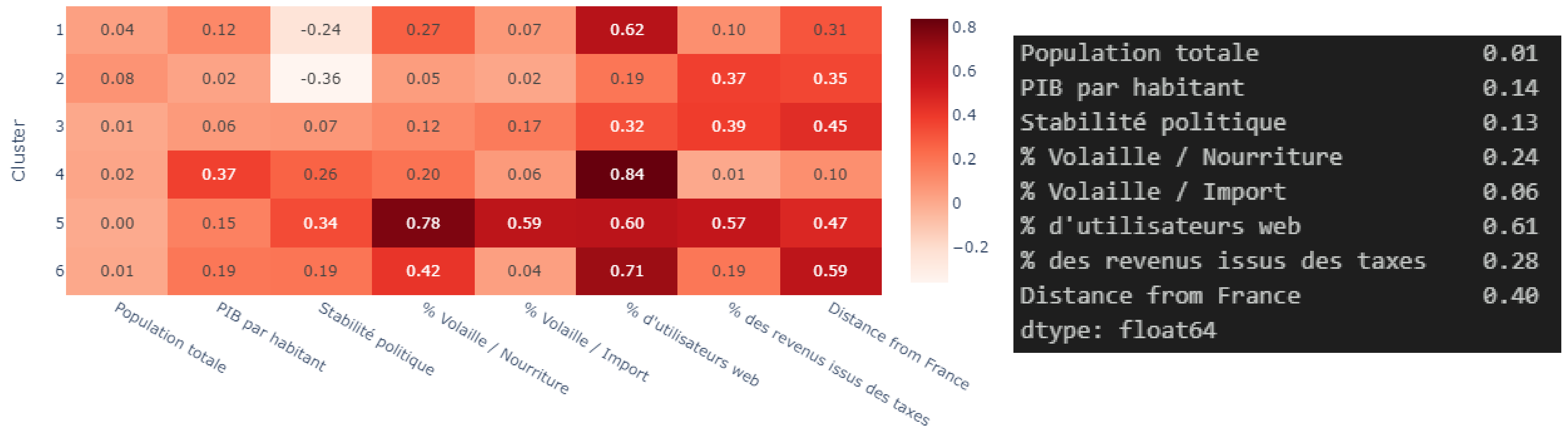


Le dendrogramme met en évidence un nb de cluster optimal entre 3 et 6.
On choisir 6, qui est le nombre de cluster maximal proposé.

Démarche : Modélisation

Etape 2 : Analyse des clusters CAH

Visualisation des centroïdes en heatmap



On compare les centroïdes des clusters avec les valeurs médianes pour chaque variable

Démarche : Modélisation

Etape 2 : Analyse des clusters CAH

- **Cluster 1** : Un certain potentiel mais une forte instabilité politique. La plupart des valeurs sont proches des médianes voire supérieures.
- **Cluster 2** : Une très forte population et une instabilité très élevée. Ce sont les pays les moins intéressants sur tous les points.
- **Cluster 3** : Pays peu peuplés, pauvres, instables et ayant des valeurs inférieures aux médianes.
- **Cluster 4** : Pays affichant le potentiel le plus intéressant sur la globalité des critères.
- **Cluster 5** : Bon potentiel car grosse conso de volaille et imports élevés mais très éloignés de la France et haut niveau de taxes.
- **Cluster 6** : Pays peu peuplés mais affichant des données intéressantes sur la consommation de volaille et les taxes. Ils importent cependant assez peu et sont surtout très éloignés de la France.

Le cluster 4 est le plus intéressant ici et sera retenu pour la suite. Bien que présentant certains points positifs, les autres seront écartés.

Résultats

Comparatif des deux méthodes

Nous avons retenu 1 cluster par méthode. Ils semblaient similaires mais qu'en est-il vraiment?

```
liste_kmeans = df_clustering_mas_kmeans.loc[df_clustering_mas_kmeans["Cluster"] ==  
.....3].index.tolist()  
liste_cah = df_clustering_mas_cah.loc[df_clustering_mas_cah["Cluster"] ==  
.....4].index.tolist()  
  
difference_cluster = list(set(liste_cah).difference(liste_kmeans))  
  
print("Il y a " + str(len(difference_cluster)) +  
....." en plus pour la CAH par rapport aux pays de liste_km.")  
print()  
print(str(sorted(list(difference_cluster))))
```

Python

Il y a 3 en plus pour la CAH par rapport aux pays de liste_km.

['Grèce', 'Roumanie', 'Serbie']

Les clusters sont identiques à l'exception de 3 pays présents uniquement dans le cluster CAH. Après analyse, ils ne représentent pas d'opportunité intéressante donc nous resterons sur le cluster kmeans.

Recommandations

Dans quels pays s'installer ?

En prenant en considération tous les éléments du cadre d'analyse PESTEL, voici les recommandations de pays où développer son activité à la suite du clustering :



Albanie



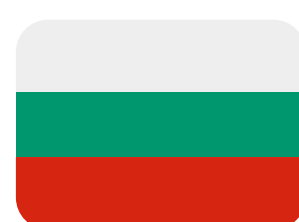
Allemagne



Autriche



Belgique



Bulgarie



Canada



Macao



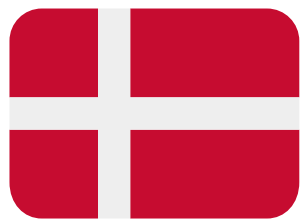
Chine



Chypre



Croatie



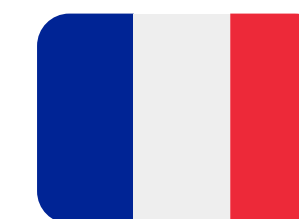
Danemark



Espagne



Finlande



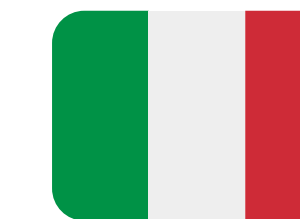
France



Hongrie



Islande



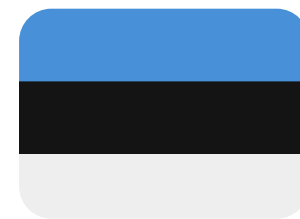
Italie



Lettonie



Irlande



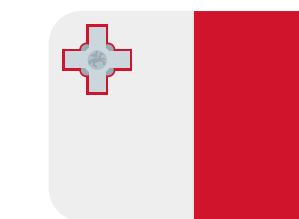
Estonie



Lituanie



Luxembourg



Malte



Norvège



Pays-Bas



Pologne



Portugal



UK



Slovaquie



Slovénie



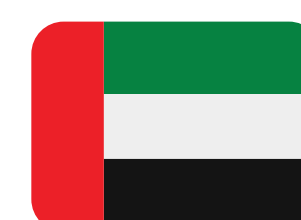
Suisse



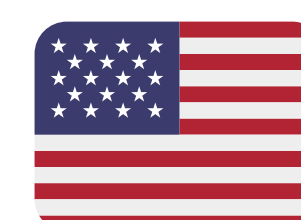
Suède



Tchéquie



EAU



États-Unis

Recommandations

En conclusion...

Chacun des pays présenté représente un marché potentiel pour l'entreprise. La sélection pourra se faire en fonction du critère qui importera le plus pour les décideurs.

Une analyse plus fine et une réduction du cluster est envisageable pour établir une shortlist de pays.

Cette analyse peut également être couplée à d'autres analyses macroéconomiques prenant en compte d'autres variables et reposant sur d'autres cadres ou bien couplée à un SWOT pour diagnostiquer la partie interne à l'entreprise.



La poule qui chante

Merci pour votre attention !