

Produisez une étude de marché avec R ou Python

Hébert Thomas

Data Analyst

21/02/2025



Contexte

- **La Poule qui Chante, entreprise spécialisée en poulet bio, veut s'ouvrir à l'international. Aucun marché n'est encore ciblé.**
- **L'objectif est d'identifier les zones clés selon le besoin, le climat institutionnel et la facilité des échanges commerciaux.**
- **Cette analyse guidera notre stratégie avant une étude approfondie des marchés les plus prometteurs.**



Sommaire

1. Récolte des données

- Sources de données
- Sélection des variables

2. Préparation et nettoyage des données

- Fusion des sources et structuration
- Gestion des valeurs manquantes et anomalies
- Transformation et standardisation

3. Analyse et segmentation des marchés

- ACP : réduction de dimensions et visualisation
- Clustering : CAH et k-means
- Interprétation des groupes et recommandations



Récolte des données



Sources de données



Organisation des Nations Unies
pour l'alimentation et l'agriculture

- Disponibilité Alimentaire (2017)
- Population (2000-2018)



- ISO Pays



WORLD BANK GROUP

- **Doing Business (2017)**
Indice de facilité de faire des affaires du avec le pays (0 à 100)
- **Political Stability (2017)**
Indice de stabilité politique du pays (-2,5 à 2,5)
- **Rule Of Law (2017)**
Indice de confiance du respect des lois du pays (-2,5 à 2,5)
- **PIB Habitant (2017)**



Sélection des variables

Population

Colonne	Type de donnée
Code zone	Code du pays (non ISO)
Zone	Nom du pays
Année	année concernée
Valeur	Population en nombre

Pré-traitement :

- année 2017

Disponibilité

Colonne	Type de donnée
Zone	Nom du pays
Production	Quantité en KT
Importation	Quantité en KT
Exportation	Quantité en KT
Disponibilité alimentaire	Quantité en kg/pers/an
Disponibilité protéines	Quantité en g/pers/j

Pré-traitement :

- pivot de la table
- volaille

ISO Pays

Colonne	Type de donnée
iso3	Code ISO du pays
name_en	Nom du pays en anglais
name_fr	Nom du pays
name_native	Nom natif du pays

PIB Habitant

Colonne	Type de donnée
Country_code	Code ISO du pays
Country	Nom du pays en anglais
PIB_hab	\$/hab

Doing Business

Colonne	Type de donnée
Country_code	Code ISO du pays
Country	Nom du pays en anglais
DB_score	Score

Rule Of Law

Colonne	Type de donnée
Country_code	Code ISO du pays
Country	Nom du pays en anglais
ROL_score	Score

Political Stability

Colonne	Type de donnée
Country_code	Code ISO du pays
Country	Nom du pays en anglais
PS_score	Score



Préparation et nettoyage des données



Fusion des sources et structuration

Population

Colonne	Type de donnée
Zone	Nom du pays

Disponibilité

Colonne	Type de donnée
Zone	Nom du pays

ISO Pays

Colonne	Type de donnée
iso3	Code ISO du pays
name_fr	Nom du pays

PIB Habitant

Colonne	Type de donnée
Country_code	Code ISO du pays

Doing Business

Colonne	Type de donnée
Country_code	Code ISO du pays

Rule Of Law

Colonne	Type de donnée
Country_code	Code ISO du pays

Political Stability

Colonne	Type de donnée
Country_code	Code ISO du pays



Fusion des sources et structuration

Df_final	
Colonne	Type de donnée
Zone	Nom du pays
Production	Quantité en KT
Importation	Quantité en KT
Exportation	Quantité en KT
Disponibilité alimentaire	Quantité en kg/pers/an
Disponibilité protéines	Quantité en g/pers/jour
Population	Population en nombre
PS_score	Score
ROL_score	Score
DB_score	Score
PIB_hab	\$/hab



Gestion des valeurs manquantes et anomalies

- **Changement des NaN par 0 dans les colonnes Production, Importations et Exportations.**
- **Application des scores PS, ROL et DB de la France à la Nouvelle-Calédonie et la Polynésie Française**
- **Application du score DB de la Chine à Macao**
- **Suppression de la Corée du Nord et Cuba (valeurs manquantes)**
- **Application de la moyenne du score DB des pays frontaliers au Turkménistan**
- **Ajout manuel du PIB du Venezuela (source Wikipédia)**



Transformation et standardisation

- **Division des colonnes Production, Importations et Exportations par la colonne Population puis multiplication par 1 000 000 pour ramener les valeurs en kg**
- **Suppression de la colonne Population**
- **Suppression de la colonne Disponibilité de protéines (forte corrélation avec Disponibilité alimentaire)**



Transformation et standardisation

Extrait du Dataframe final (avant standardisation des données)

Zone	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule Of Law	Doing Business
Afghanistan	0.771432	0.798984	0.000000	1.53	525.469771	-2.79	-1.58	39.2
Afrique du Sud	29.240609	9.016001	1.105074	35.69	6618.335083	-0.28	-0.16	64.9
Albanie	4.507364	13.175372	0.000000	16.36	4531.032207	0.37	-0.41	64.0
Algérie	6.644247	0.048322	0.000000	6.38	4554.667540	-0.92	-0.93	45.8
Allemagne	18.316346	10.186501	7.815297	19.47	45526.599958	0.57	1.57	79.4

167 pays sont présents pour l'analyse

La standardisation consiste à centrer et réduire les données afin qu'une variable ne pèse pas plus qu'une autre dans la balance (Exemple : le PIB qui peut atteindre des dizaines de milliers comparé aux autres variables).



Analyse et segmentation des marchés



ACP : réduction de dimensions et visualisation

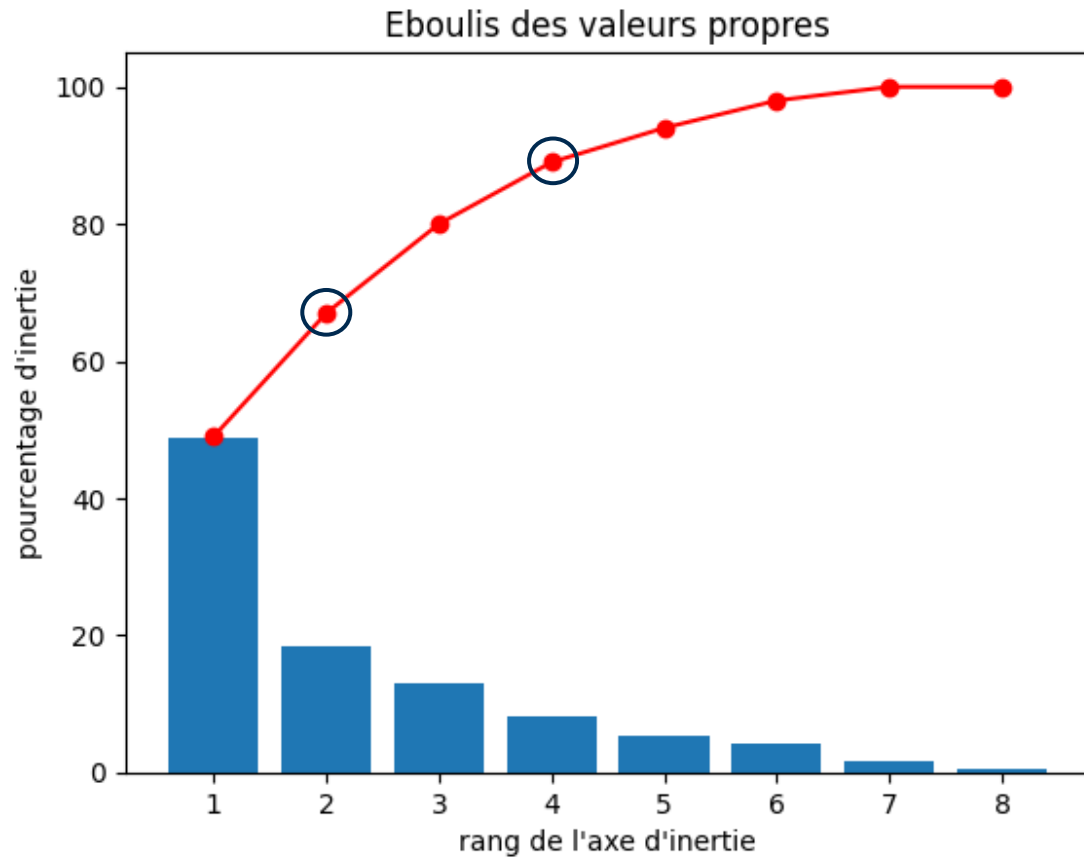
Variable	Composante
Disponibilité alimentaire	F1
Political Stability	
Rule Of Law	
PIB	
Doing Business	
Importations	F2
Exportations	
Production	

L'Analyse en Composantes Principales (ACP) permet de réduire la dimension des données tout en conservant un maximum d'informations.

Les variables corrélées sont synthétisées en composantes principales.



ACP : réduction de dimensions et visualisation

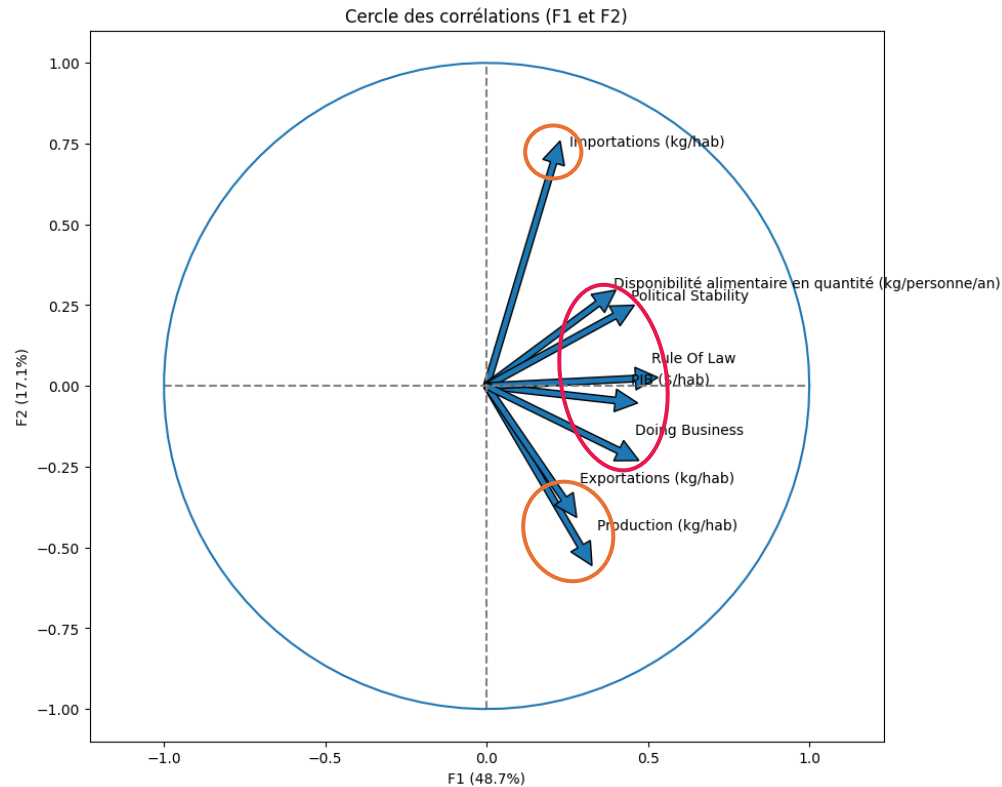


Les deux premières composantes principales, F1 et F2, expliquent 66% de l'inertie totale.

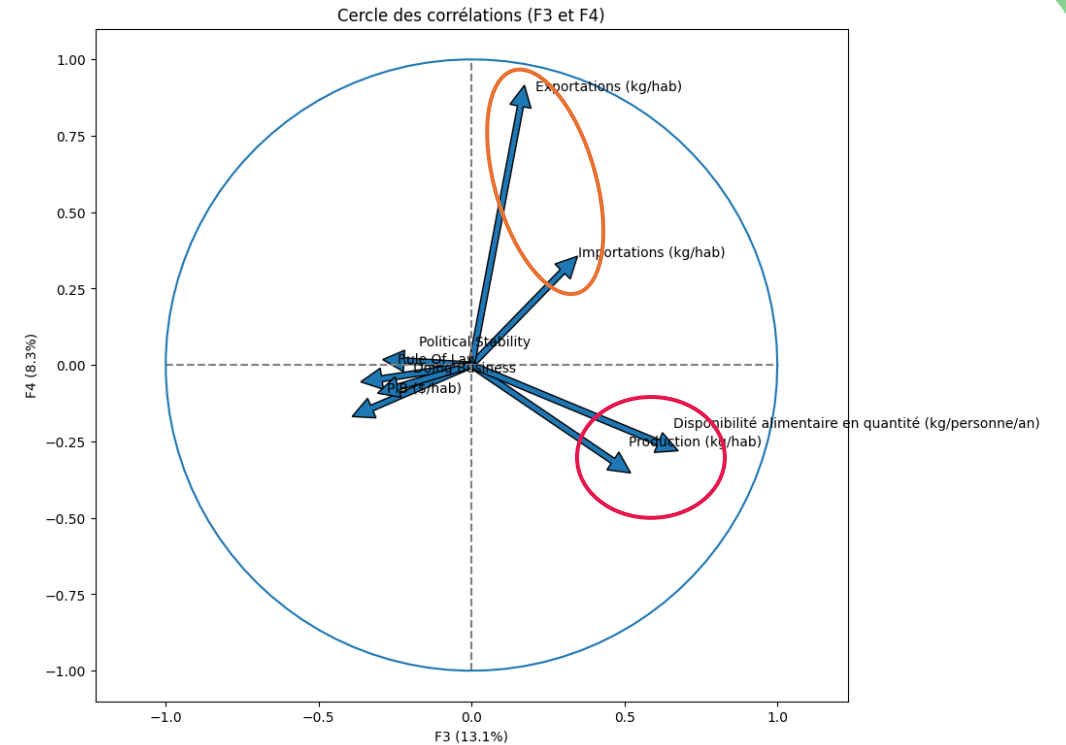
L'ajout des composantes F3 et F4 permet d'atteindre 90% de l'inertie.



ACP : réduction de dimensions et visualisation

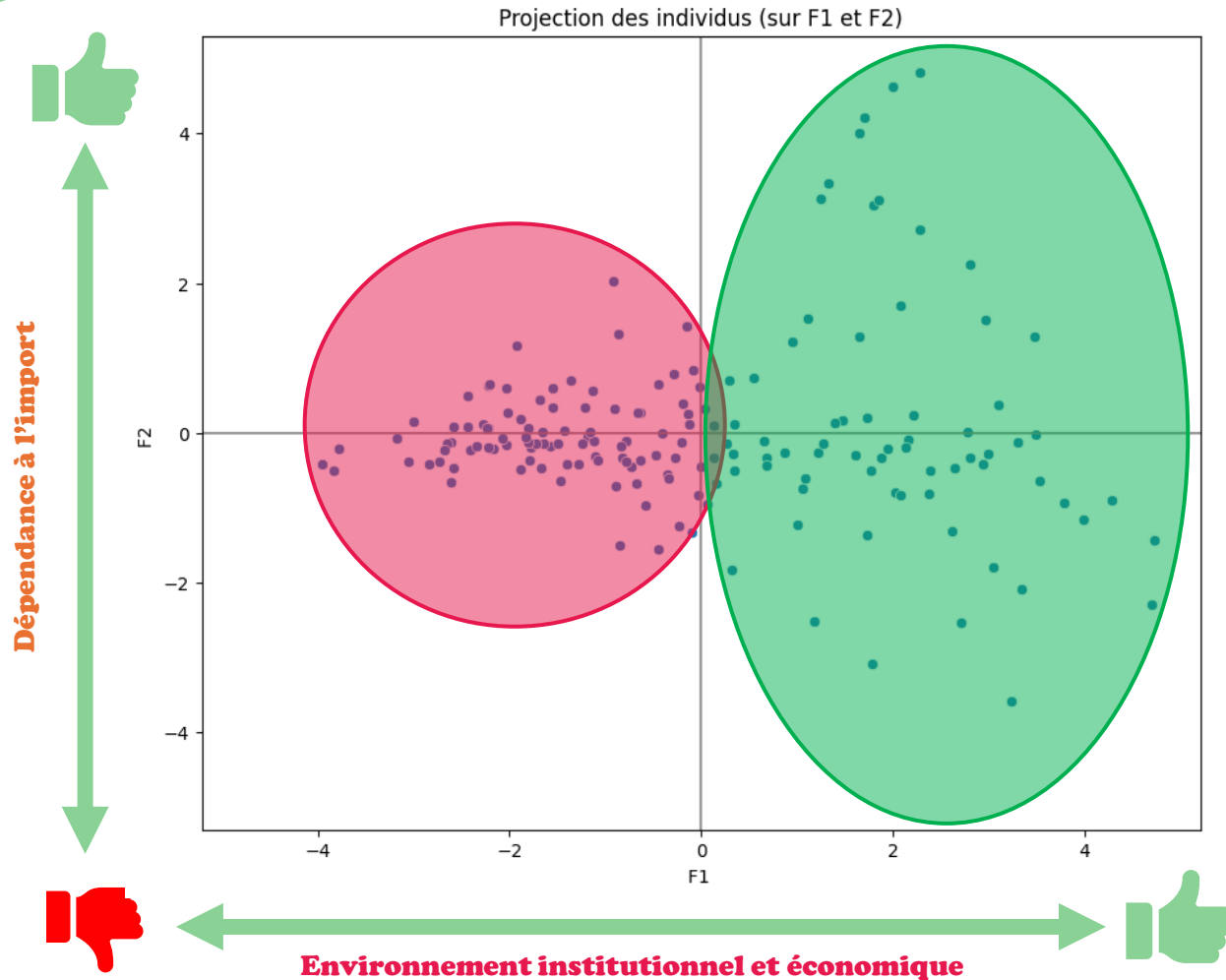


**F1 = Environnement
institutionnel et économique**
F2 = dépendance à l'import



F3 = indépendance alimentaire
F4 = échanges commerciaux

ACP : réduction de dimensions et visualisation

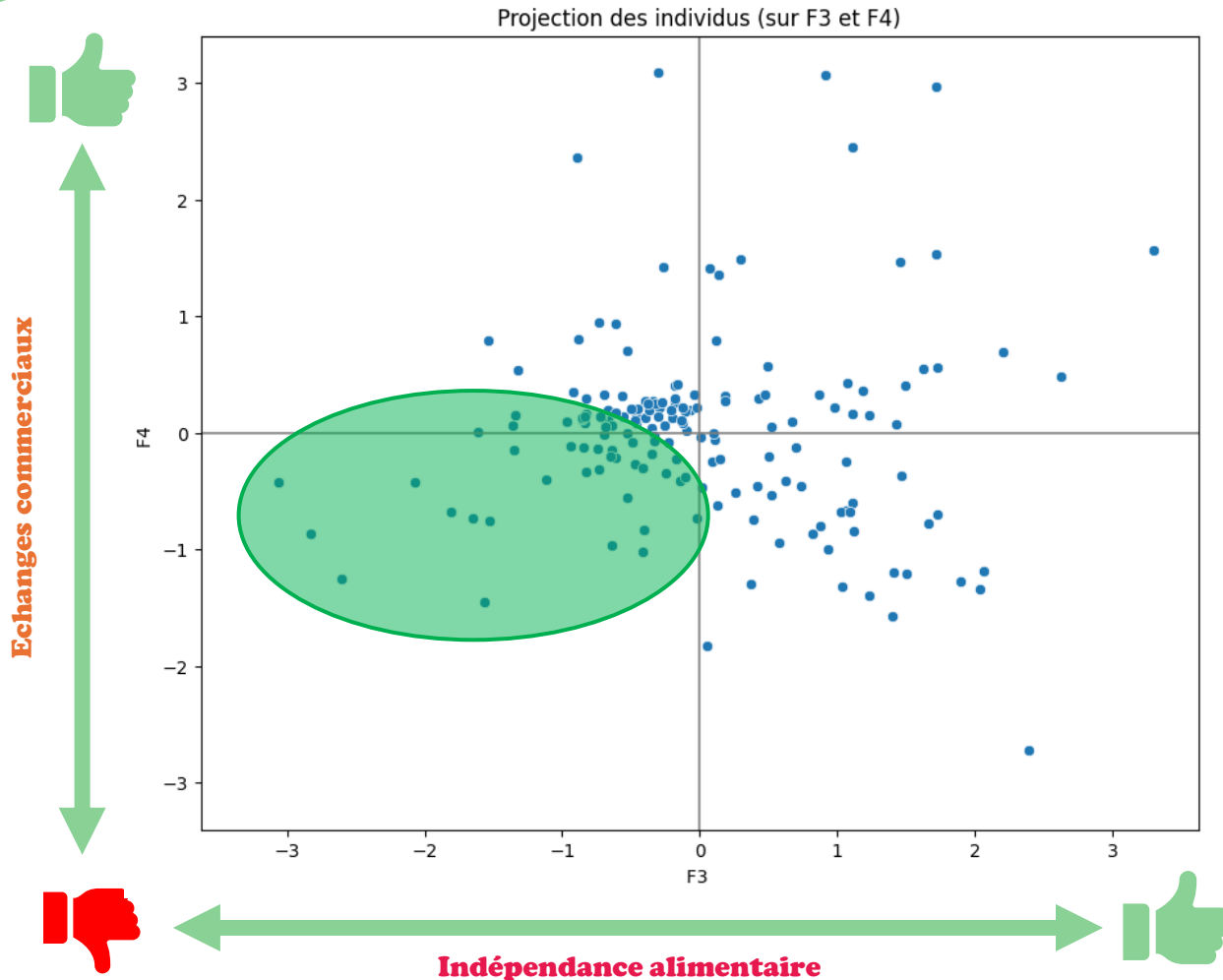


En projetant nos variables sur le premier plan factoriel (**F1** et **F2**), nous identifions un groupe (indiqué par la zone rouge), qui semble hors scope de notre cible car **trop instable sur les plans institutionnel et économique**.

F2 ne nous permet pas concrètement de définir une cible, un pays **dépendant de l'import** peut ne pas représenter un marché pour autant.

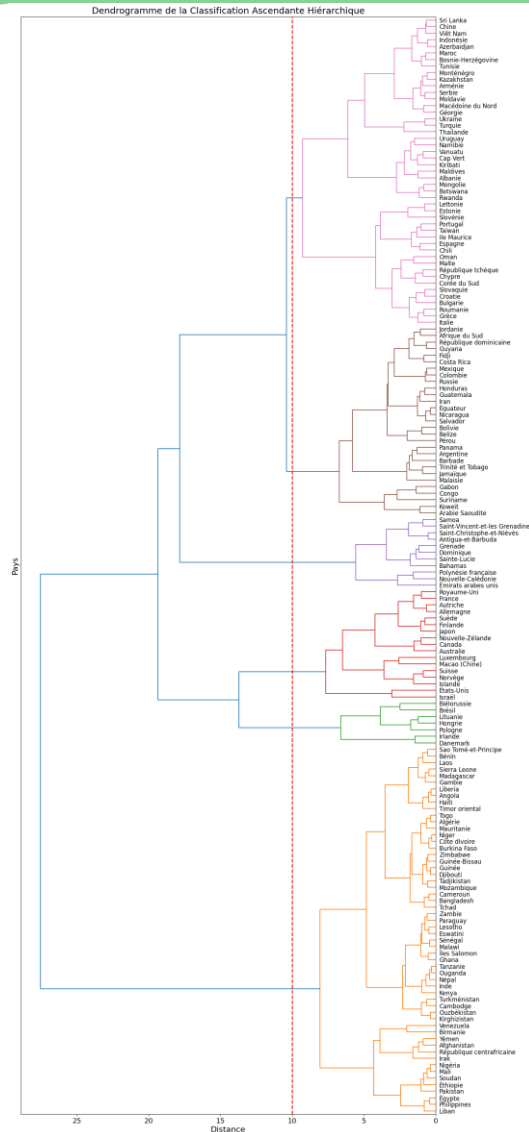
Tous les individus de la zone verte peuvent donc être une cible à ce moment de l'analyse.

ACP : réduction de dimensions et visualisation



En projetant nos variables sur le second plan factoriel (**F3** et **F4**), nous identifions un groupe (indiqué par le cercle vert) qui semble être notre cible car en **déficit en termes d'indépendance alimentaire** et pratiquant **peu d'échange commerciaux**.

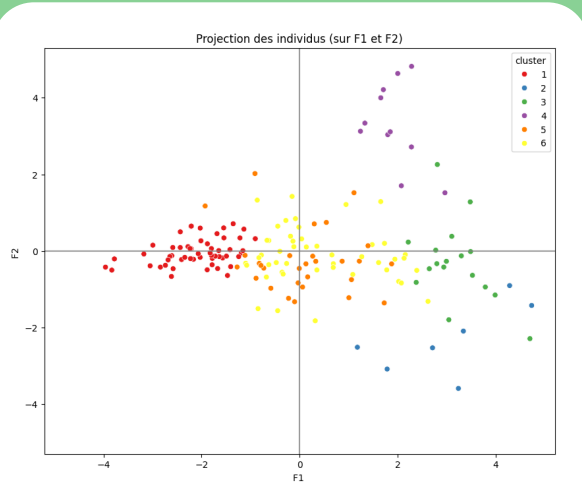
Clustering : CAH



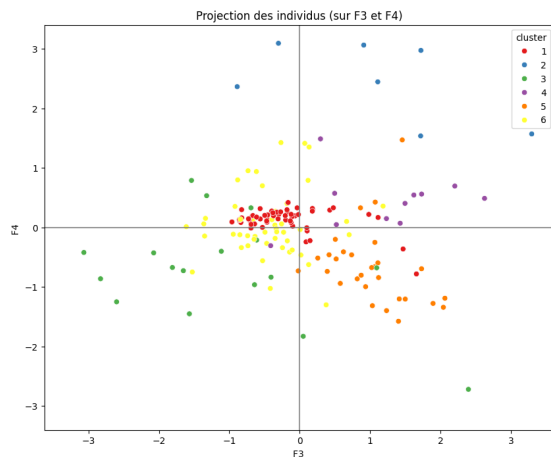
Le découpage en 6 clusters est pertinent, car il garantit une bonne répartition des pays pour l'analyse.

De plus, il assure une cohérence avec le nombre de clusters de K-Means, facilitant ainsi la comparaison.

Clustering : CAH



F1 environnement institutionnel et économique
F2 dépendance à l'import

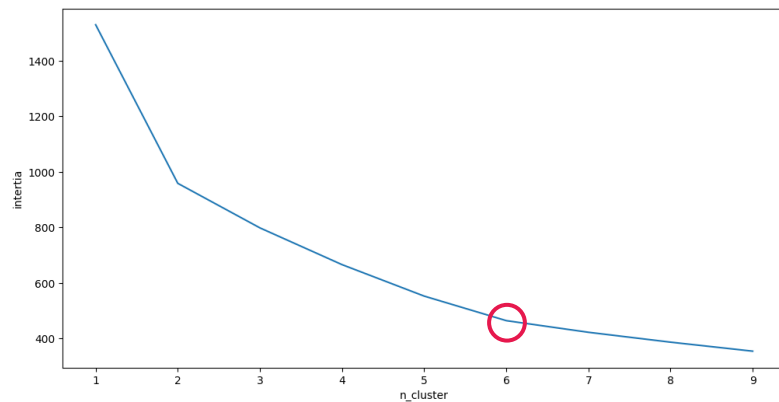


F3 indépendance alimentaire
F4 échanges commerciaux

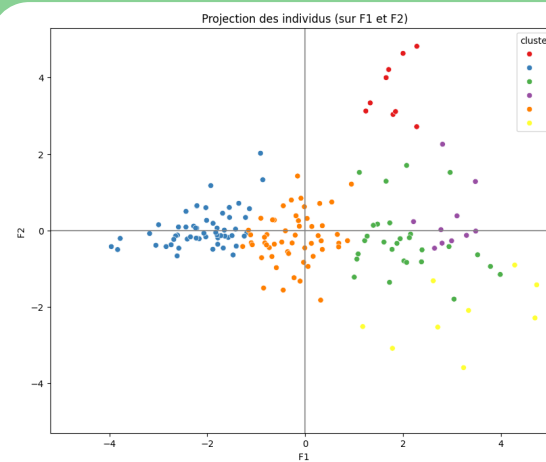
	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule Of Law	Doing Business
cluster								
1	3.686722	2.545050	0.022857	5.770893	2001.698043	-0.818929	-0.854107	48.889286
2	46.389714	9.877140	21.815097	30.652857	27329.458793	0.482857	0.555714	74.285714
3	28.928501	8.001019	3.486125	29.907059	57160.515150	0.866471	1.625294	78.464706
4	3.868350	62.624067	0.900735	50.661818	18840.345224	0.785455	0.698182	63.118182
5	30.469960	8.190009	0.766886	32.170000	8979.662317	-0.129310	-0.389310	59.803448
6	14.525028	8.131635	2.465707	17.627660	11614.572486	0.191277	0.245106	69.402128

Les clusters appliqués aux différentes projections indiquent que le cluster 3 (en vert sur les graphiques) est un choix idéal pour notre cible de pays, confirmé par le tableau des moyennes des valeurs.

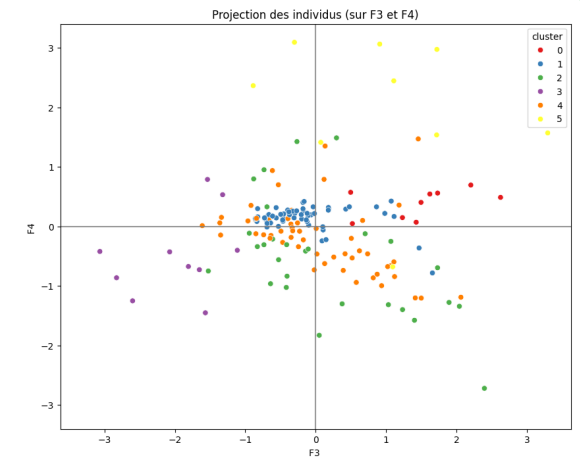
Clustering : k-means



Pour définir le nombre de clusters nous appliquons la méthode du coude et choisissons ici 6 clusters.



**F1 environnement institutionnel et économique
F2 dépendance à l'import**

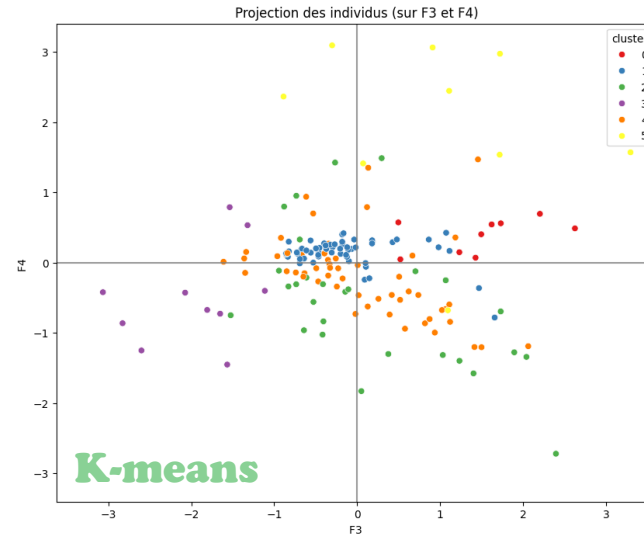
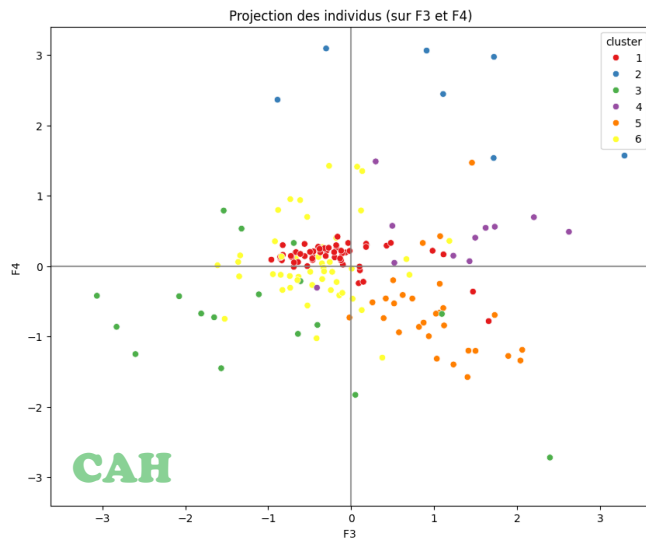
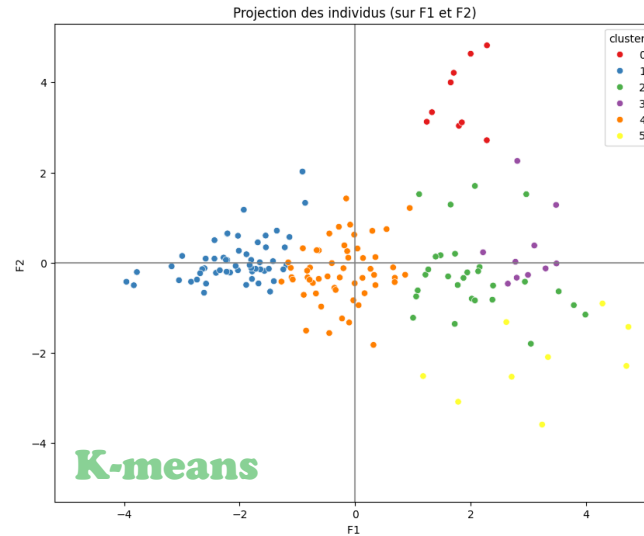
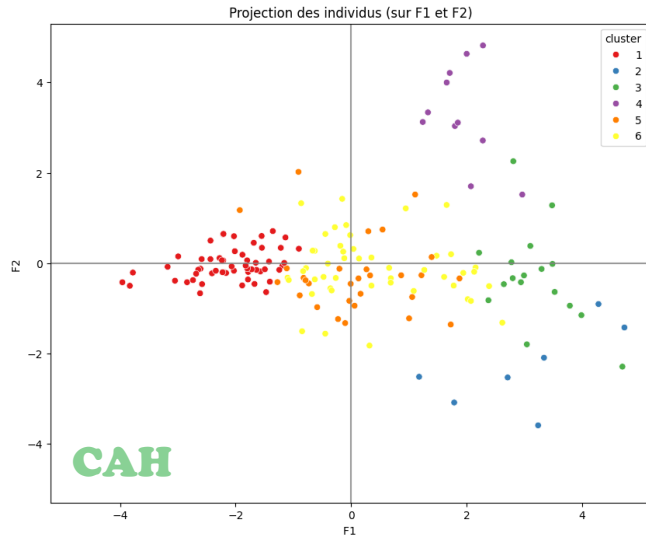


**F3 indépendance alimentaire
F4 échanges commerciaux**

	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule Of Law	Doing Business
cluster								
0	3.764917	67.861214	0.000000	52.788889	14727.610607	0.863333	0.612222	60.077778
1	3.592906	3.823791	0.021415	6.611754	2249.853509	-0.815789	-0.878070	48.305263
2	31.575968	10.889321	3.170952	34.788571	26605.534450	0.546786	0.862143	73.907143
3	15.628087	11.034279	2.339926	20.905000	64920.477136	1.118000	1.669000	76.830000
4	17.146750	6.169272	1.293518	19.168704	6644.213684	-0.113704	-0.207037	64.281481
5	47.423736	8.794492	19.781027	32.740000	30547.950076	0.501111	0.721111	75.466667

Il y a beaucoup de similitudes avec CAH. Le cluster 3 (en violet ici) correspond à notre cible.

Clustering : CAH vs k-means



Sur le premier plan factoriel (F1 et F2), les clusters obtenus avec K-means sont beaucoup plus compacts et regroupées que ceux générés par la méthode CAH.

En revanche, sur le deuxième plan factoriel (F3 et F4), la différence entre les deux méthodes est moins marquée.

Interprétation des groupes et recommandations

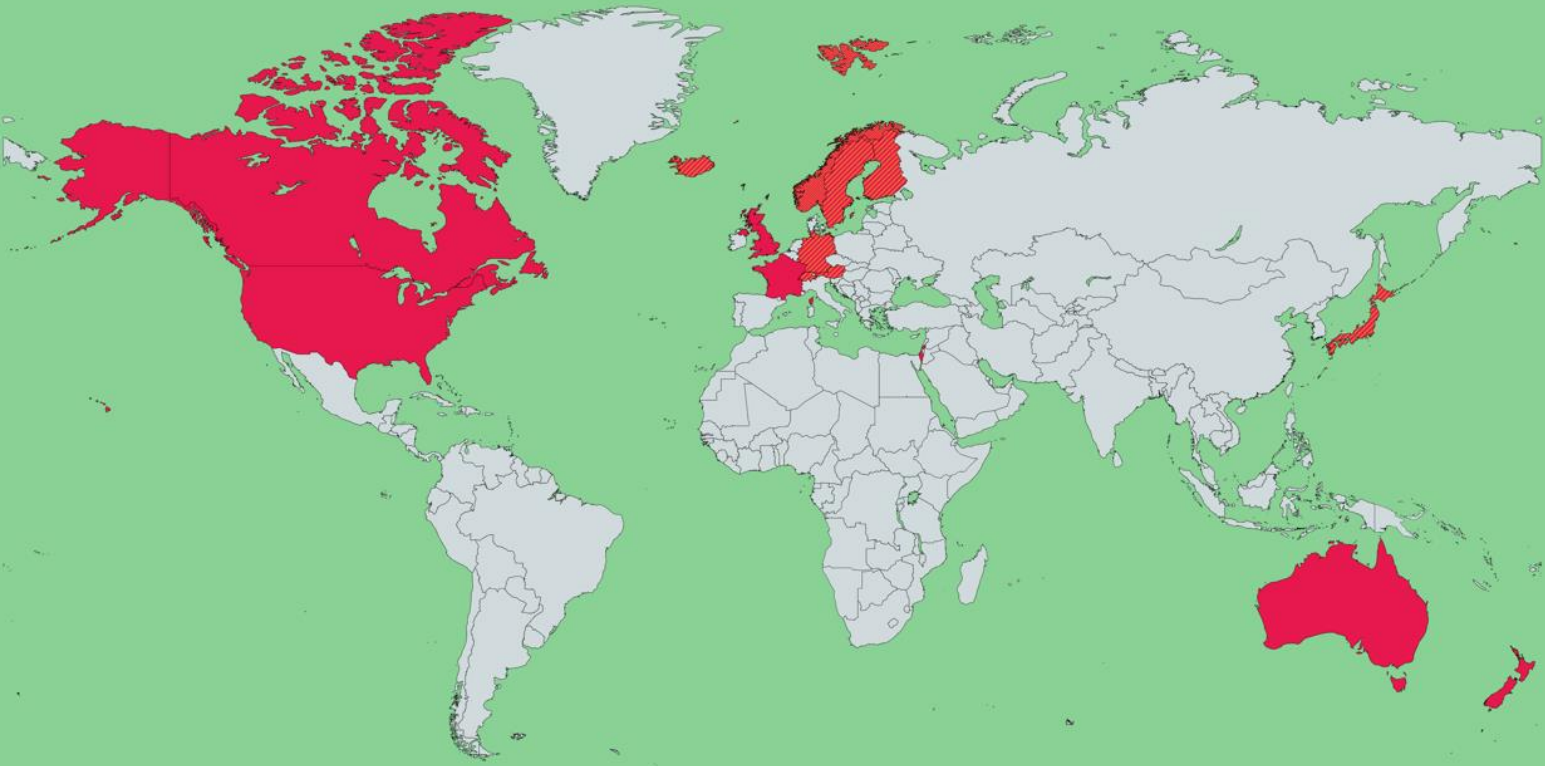
CAH cluser 3

	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule of Law	Doing Business	Cluster
Zone									
Allemagne	10.316346	10.186501	7.815297	19.47	45526.599958	0.57	1.57	79.4	3
Australie	51.817637	0.650813	1.708385	47.65	53901.714539	0.88	1.65	80.2	3
Autriche	16.780234	12.471795	8.843637	18.20	47163.742578	1.03	1.80	78.5	3
Canada	38.576818	4.954795	4.437536	39.02	45128.628117	1.09	1.76	80.0	3
Macao (Chine)	4.818619	36.942747	0.000000	34.47	77704.921531	1.41	0.81	63.8	3
Finlande	23.406154	2.903089	2.177317	18.76	46085.017474	1.06	2.02	79.9	3
France	26.985468	7.803523	7.726413	22.90	30687.162641	0.27	1.40	76.2	3
Irlande	29.904932	5.980986	0.000000	29.95	72010.149032	1.34	1.57	78.9	3
Israël	76.299320	0.000000	0.363908	67.39	40997.911024	-0.89	0.99	75.4	3
Japon	17.372178	8.384135	0.078430	18.50	30834.052934	1.10	1.53	78.3	3
Luxembourg	0.000000	18.583906	1.689446	18.33	110193.213797	1.31	1.70	69.1	3
Norvège	19.069823	0.377620	0.000000	19.05	76131.838403	1.15	1.98	82.1	3
Nouvelle-Zélande	47.426284	0.212674	4.891500	34.79	42913.188956	1.56	1.88	86.8	3
Royaume-Uni	27.185210	11.674354	5.380094	31.94	40572.121482	0.38	1.65	83.3	3
Suisse	10.761839	6.031360	0.473048	15.72	82254.378827	1.24	1.89	76.1	3
Soudé	15.850747	8.480654	2.322084	16.60	53300.858728	0.97	1.82	82.2	3
États-Unis	67.410113	0.378363	11.357038	55.68	60322.261424	0.26	1.61	83.7	3

K-means cluster 3

	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule of Law	Doing Business	Cluster
Zone									
Allemagne	10.316346	10.186501	7.815297	19.47	45526.599958	0.57	1.57	79.4	3
Autriche	16.780234	12.471795	8.843637	18.20	47163.742578	1.03	1.80	78.5	3
Macao (Chine)	4.818619	36.942747	0.000000	34.47	77704.921531	1.41	0.81	63.8	3
Finlande	23.406154	2.903089	2.177317	18.76	46085.017474	1.06	2.02	79.9	3
Irlande	29.904932	5.980986	0.000000	29.95	72010.149032	1.34	1.57	78.9	3
Japon	17.372178	8.384135	0.078430	18.50	30834.052934	1.10	1.53	78.3	3
Luxembourg	0.000000	18.583906	1.689446	18.33	110193.213797	1.31	1.70	69.1	3
Norvège	19.069823	0.377620	0.000000	19.05	76131.838403	1.15	1.98	82.1	3
Suisse	10.761839	6.031360	0.473048	15.72	82254.378827	1.24	1.89	76.1	3
Soudé	15.850747	8.480654	2.322084	16.60	53300.858728	0.97	1.82	82.2	3

Un total de 17 pays a été extrait par la méthode CAH, tandis que 10 pays ont été identifiés par K-means. Sur la carte, on peut observer que ces pays sont répartis à travers le monde.



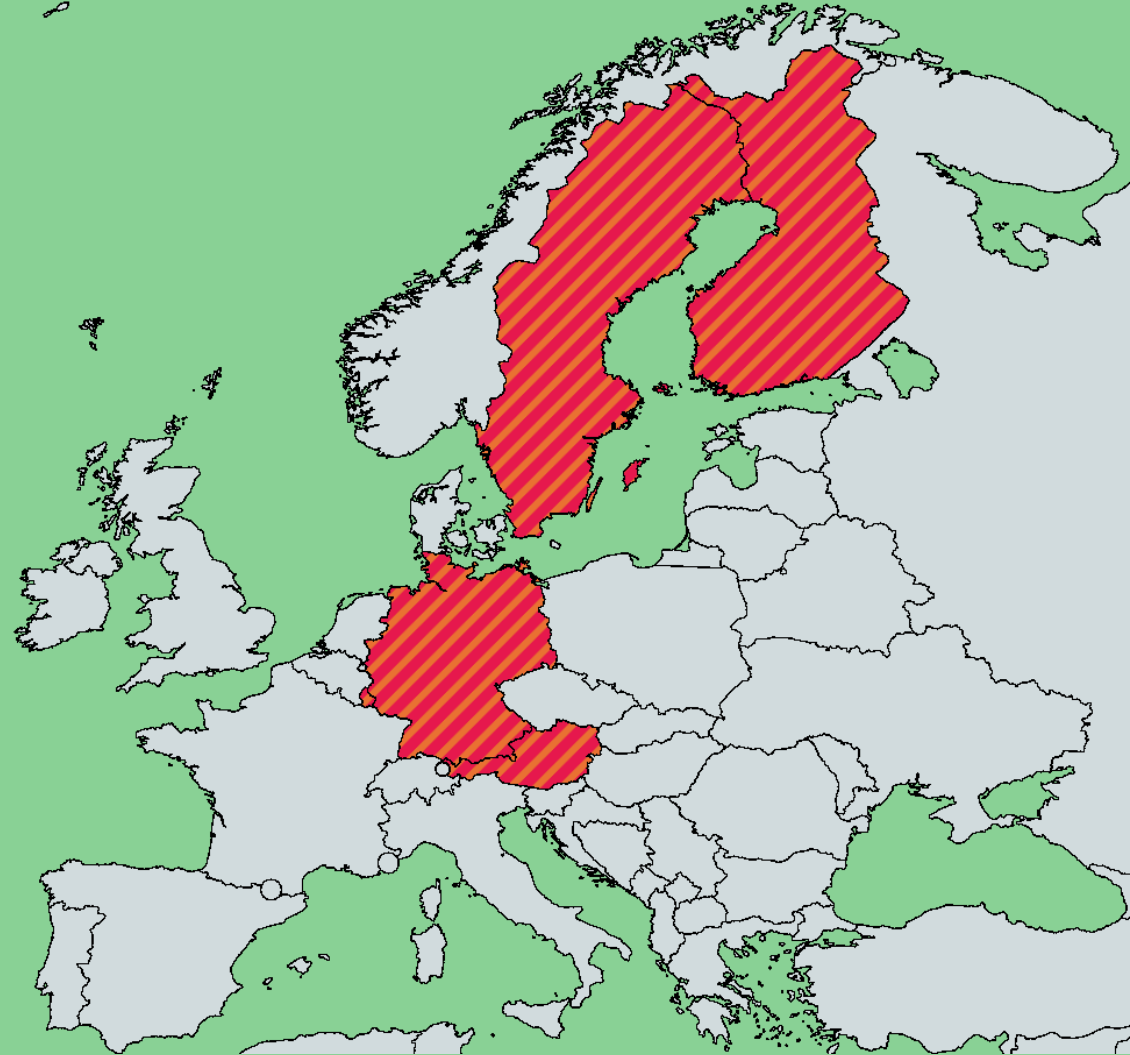
Pour optimiser les aspects opérationnelles, stratégiques et logistiques, il est préférable de recentrer la cible sur les pays de l'Union Européenne.



Interprétation des groupes et recommandations

Zone	Production (kg/hab)	Importations (kg/hab)	Exportations (kg/hab)	Disponibilité alimentaire en quantité (kg/personne/an)	PIB (\$/hab)	Political Stability	Rule Of Law	Doing Business	cluster
Allemagne	18.316346	10.186501	7.815297	19.47	45526.599958	0.57	1.57	79.4	3
Autriche	16.780234	12.471795	8.843637	18.20	47163.742578	1.03	1.80	78.5	3
Finlande	23.406154	2.903089	2.177317	18.76	46085.017474	1.06	2.02	79.9	3
Luxembourg	0.000000	18.583906	1.689446	18.33	110193.213797	1.31	1.70	69.1	3
Suède	15.850747	8.480654	2.322084	16.60	53300.858728	0.97	1.82	82.2	3

Nous avons identifié cinq pays qui représentent un marché prometteur pour l'exportation de notre poulet. Ces pays se caractérisent par des conditions politico-économiques favorables et un déficit en termes de disponibilité alimentaire, créant ainsi une opportunité pour répondre à leurs besoins. De plus, le fait que la France se trouve dans le même cluster d'origine (CAH) que ces pays renforce notre confiance dans la pertinence de ce choix, suggérant un positionnement stratégique favorable à l'exportation



**Merci pour votre
attention**

