Étude de marché

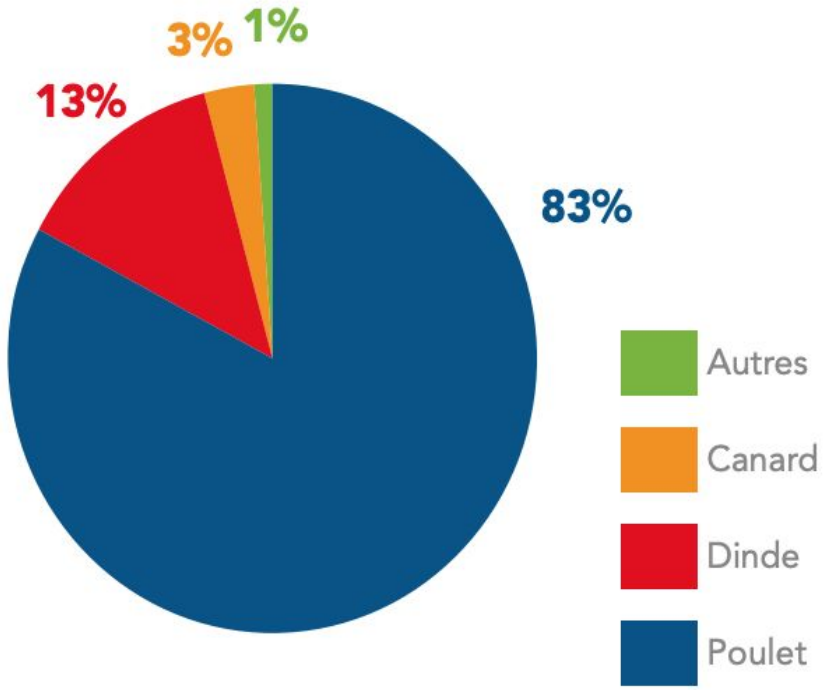
Identifier des pays propices à une insertion dans le marché du poulet

Projet 5 - Nalron Novembre 2019
OpenClassrooms - ENSAE-ENSAI Formation Continue

Le marché du Poulet…

Répartition de la production de la volaille par espèce.

Source : CIRCABC



CIRCABC: Communication and Information Resource Centre for Administrations, Businesses and Citizens.

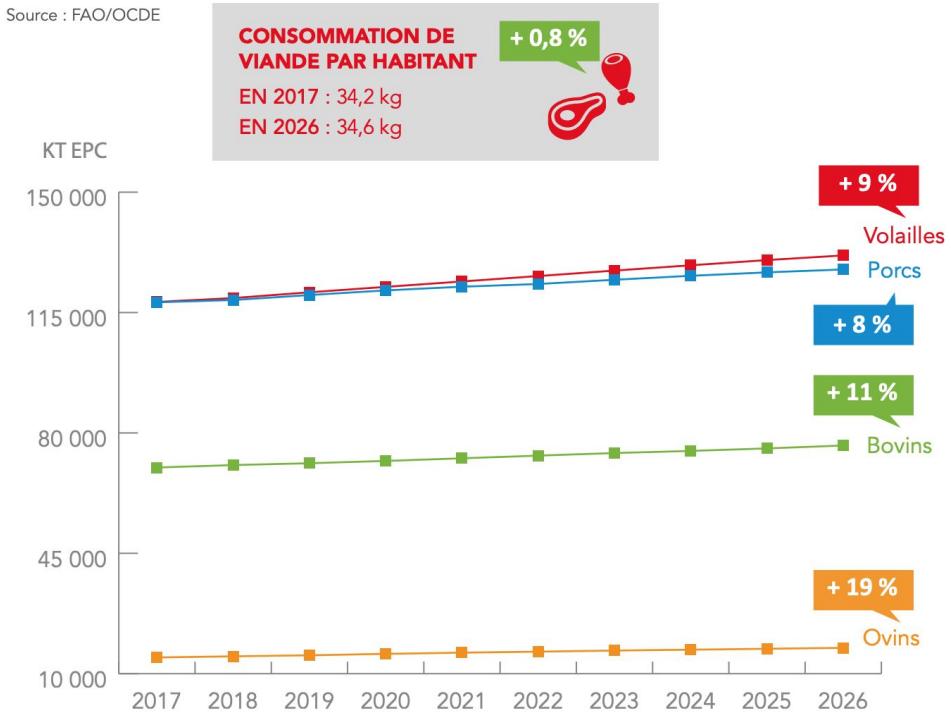
Prospective

Prévisions d'évolution de la demande mondiale à 2026.

Le marché est prometteur, un plan de développement d'exportation peut-être fait jusqu'à 2026 minimum…

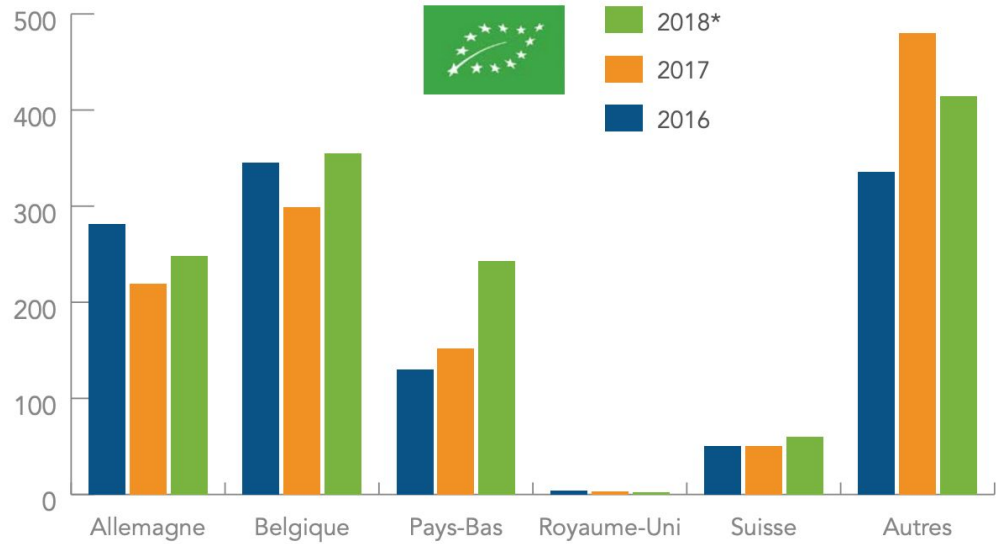
FAO/OCDE : Organisation pour l'alimentation et l'agriculture / Organisation de coopération et de développement économiques.

Source : FAO/OCDE



Tendances du marché

Source : Synalaf



Les exportations de volailles Bio croissent de 10 % en 2018 par rapport à 2017. Les découpes Bio représentent un peu plus de la moitié (56 %) de l'ensemble des volailles Bio exportées.

Les principaux pays destinataires des volailles fermières Label Rouge et biologiques restent la Belgique, l'Allemagne et les Pays-Bas. Ces trois pays représentent à eux seuls environ 58 % et 64 % des volumes exportés en Label Rouge et en Bio respectivement, le reste des exportations étant essentiellement tourné vers les autres pays de l'Union européenne.

ÉVOLUTION 2018/2017
TOTAL EN TONNES

EN 2016
1 141

EN 2017
1 205

+ 10 %

*Estimations

Synalaf : Syndicat National des Labels Avicoles de France.

Quels sont les enjeux?

- ➔ **Les enjeux du développement à l'international** sont l'équilibre du régime alimentaire et une limitation des contraintes du marché local.
- ➔ **Les enjeux sont de taille :** alimenter une population humaine avec notre savoir-faire «*Made in France*».

Quels sont les objectifs?

- ➔ **Les objectifs de l'entreprise** semblent se diriger davantage dans l'exportation et non dans une implantation et production locale.
- ➔ **Les objectifs premiers** se limitent à cibler certains pays dans le but d'approfondir ensuite l'étude de marché.

Démarche et méthodologie

Les étapes annoncées sont dépendantes du besoin de traitement attendu, de l'approche +/- fine des conclusions…

Le Traitement des données a été réalisé en **langage Python** sur un support Jupyter Notebook distribution Anaconda.

- ➔ Récupération des données, avec cycle de traitement, nettoyage, re-traitement: *échantillon*.
- ➔ Classification non supervisée (CAH et K-Means).
- ➔ Visualisation des clusters par ACP.
- ➔ Tests statistiques.

Construction d’un échantillon de travail :

*Un premier lot de variables quantitatives pour caractériser **le régime alimentaire des pays**.*

- disponibilité alimentaire en calories par habitant
- disponibilité alimentaire en protéines par habitant
- ratio protéines animales / protéines totales
- différence de population entre 2013 et 2012

*Un autre lot de variables quantitatives pour caractériser **la dimension marché des pays**.*

- PIB par habitant
- Import de poulets vivants
- Elevage de poulets

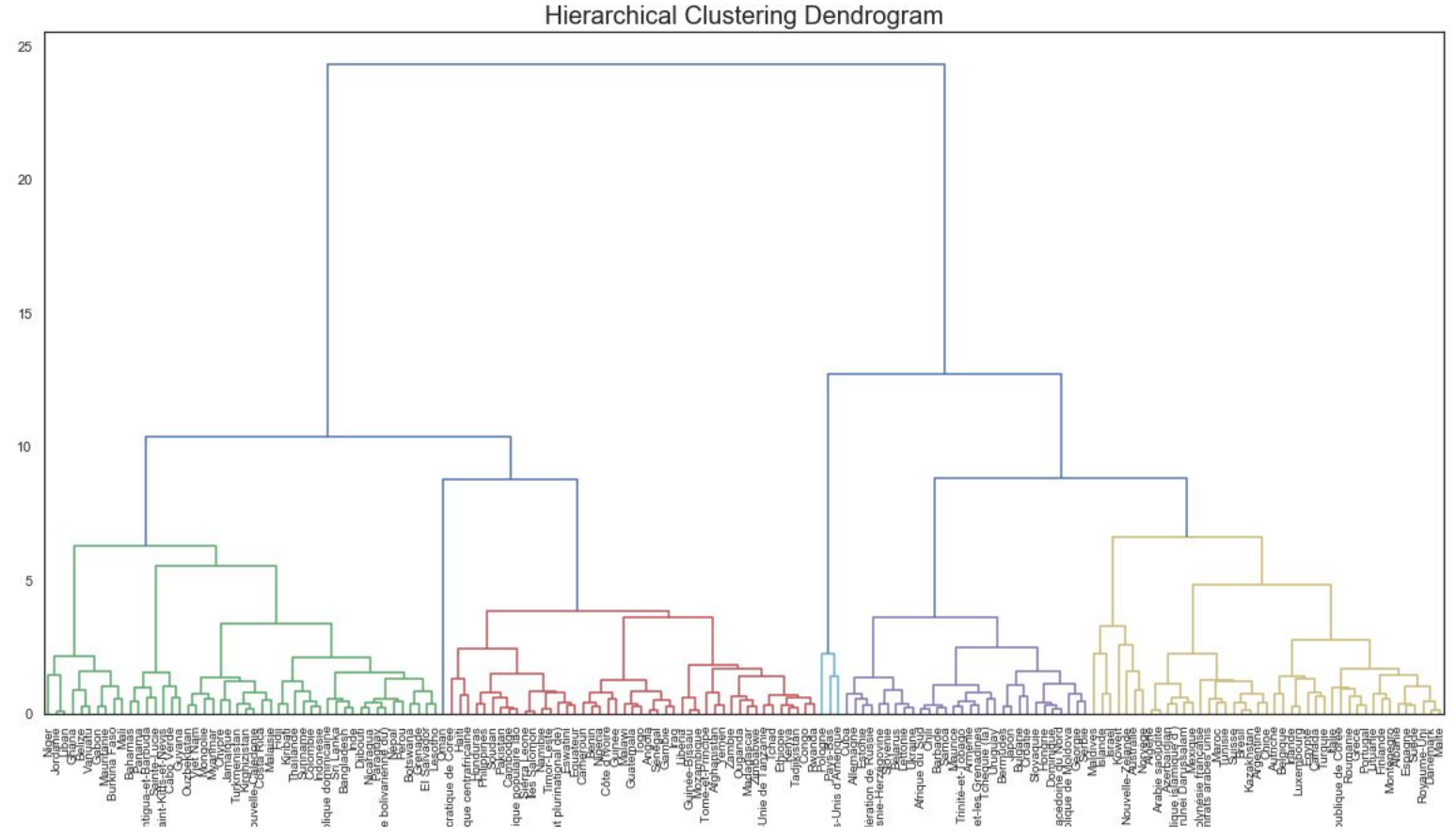
Données extraites du site fao.org sous fichiers .csv

Partitionnement - Classification hiérarchique

Premier découpage en 5 groupes, un dendrogramme a été fait pour pouvoir identifier les pays les plus similaires entre eux.

La taille de l'échantillon permet la réalisation du dendrogramme, bien que l'algorithme ait une une forte *complexité algorithmique* en temps et en espace.

```
#Création d'une Matrice des liens selon la Méthode de Ward
Z = linkage(X_scaled, method = 'ward', metric='euclidean')
```

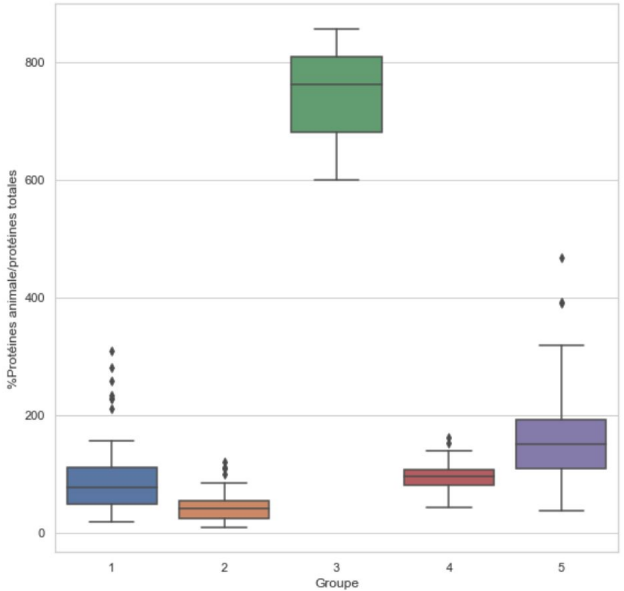

Hierarchical Clustering Dendrogram
25
20
15
10
5
0

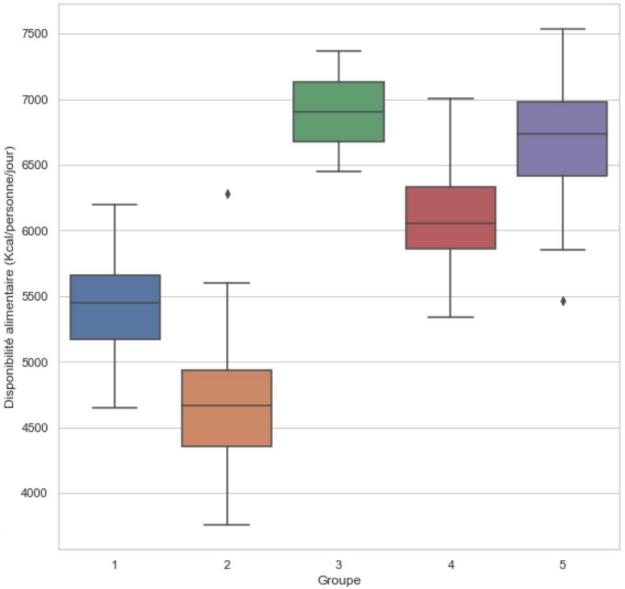
Groupe	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	%Protéines animale/protéines totales	%Evolution Population 2012/2013
1	5416.895833	149.140625	96.689792	1.475833
2	4687.086957	117.491304	44.692391	2.547826
3	6904.666667	215.043333	739.576667	0.363333
4	6072.800000	174.721667	96.508667	-0.022333
5	6680.604651	209.427907	166.132558	0.927907

Caractéristiques des 5 groupes issus du CAH.

➔ Moyennes variables
➔ Boxplot *

**Visualisation des groupes selon les variables. Ici un aperçu selon 2 variables.*





Étape transitoire vers la méthode des K-Means

46 pays sont susceptibles de devenir une cible pertinente pour l'entreprise

- ➔ Volonté d’affiner le choix vers une liste plus petite.
- ➔ Opportunité de pouvoir comparer deux méthodes.
- ➔ Plus simple, plus flexible, plus efficace…
- ➔ Facilité d’interprétation des clusters sous une forme minimisée.

Caractérisation des clusters par les centroïdes K-Means

La comparaison des clusters devient simple et rapide. Les centroïdes ci-dessous peuvent-être comparés en fonction des objectifs établis.

```
#Tableau des Centroïdes 5 clusters dans sa version centrée réduite
#La comparaison est tout de suite simplifiée, les dimensions prenant la même importance!
centroids = cls5.cluster_centers_
pd.DataFrame(centroids, columns=df_alim.columns)
```

	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	%Protéines animale/protéines totales	%Evolution Population 2012/2013
0	-0.976381	-1.000168	-0.574118	0.793794
1	1.360533	1.385893	4.075729	-0.141640
2	1.122046	1.138384	0.242587	-0.601906
3	-0.075015	-0.060516	-0.029774	-0.445590
4	0.687270	0.380713	-0.013873	6.621178

➔ *Les clusters 2 et 1 se distinguent par des moyennes plus importantes.*

5 clusters K-Means

Projection sur le premier plan factoriel par ACP.

82% de la variance expliquée par les deux premières composantes principales.

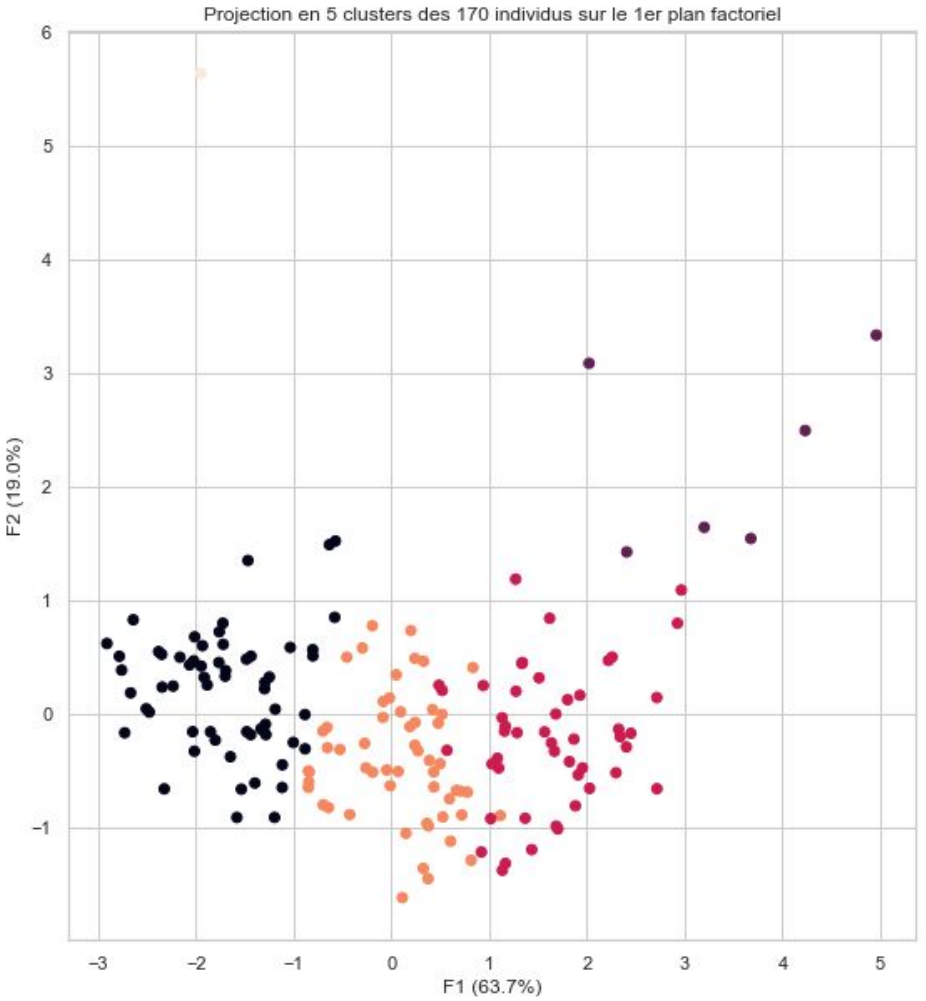
```
: #Calcul des composantes principales
  pca = decomposition.PCA(svd_solver='full')
  pca.fit(X_scaled)
```

```
PCA(copy=True, iterated_power='auto', n_components=None, random_state=None,
    svd_solver='full', tol=0.0, whiten=False)
```

```
: #Pourcentage de variance expliquée par les composantes principales à l'aide
  print(pca.explained_variance_ratio_.cumsum())
```

```
[0.63746575 0.82768495 0.970698   1.         ]
```

-> **Plus de 82% de la variance** des données est expliquée par ces deux premières composantes.



Une sélection plus fine est nécessaire

53 pays sont potentiellement intéressants suite aux deux méthodes de classification non supervisées. Avec intégration des nouvelles variables liées à la dimension marché.

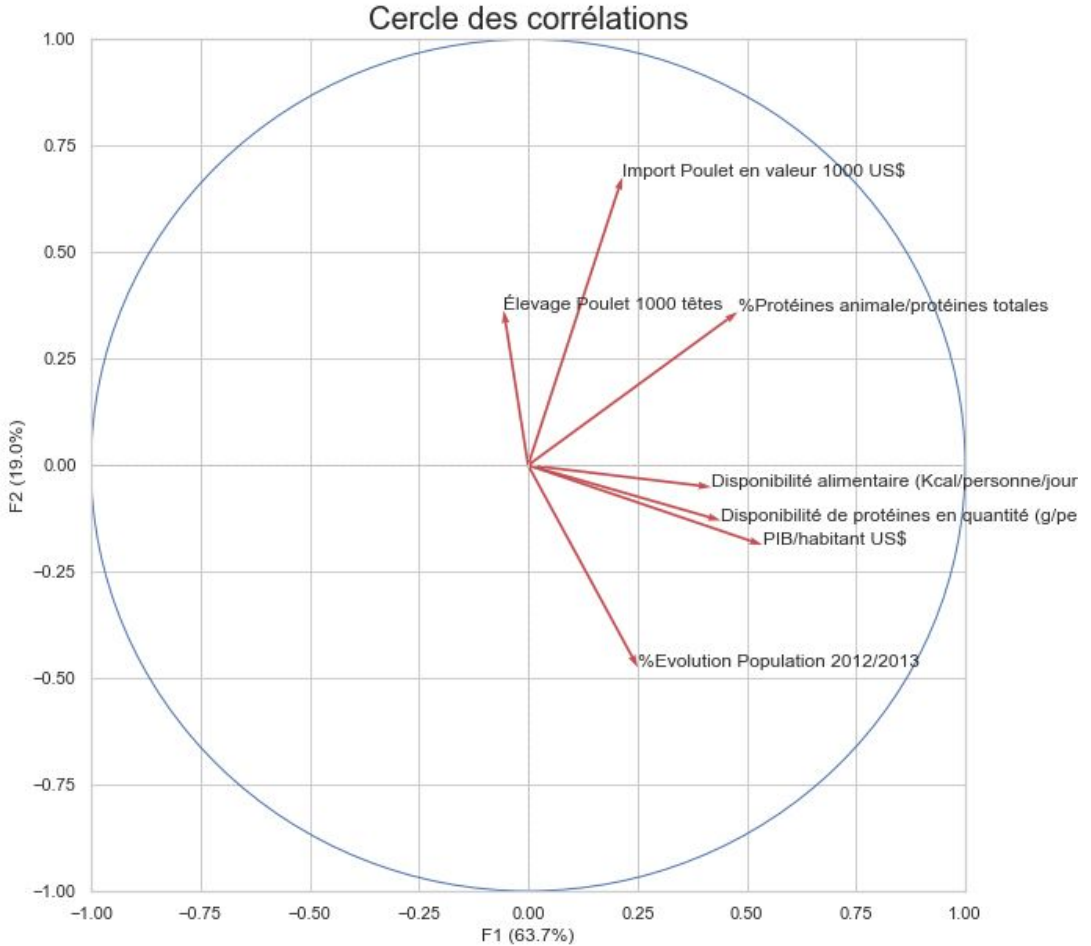
- ➔ Enrichissement de l'échantillon par les variables.
- ➔ Nouveau clustering K-Means.
- ➔ Détermination du k cluster adéquat.
- ➔ Analyse et caractérisation des clusters identifiés.

Projection des variables sur le premier plan factoriel ACP.

➔ Nouvelle base orthonormée.
➔ Variance * maximale 82%.

ACP permet une représentation en 2 dimensions, établir 2 types de profils des pays, ainsi que des corrélations entre variables.

**pca.explained_variance_ratio_ nous donne le pourcentage de variance expliquée par chacune des composantes.*

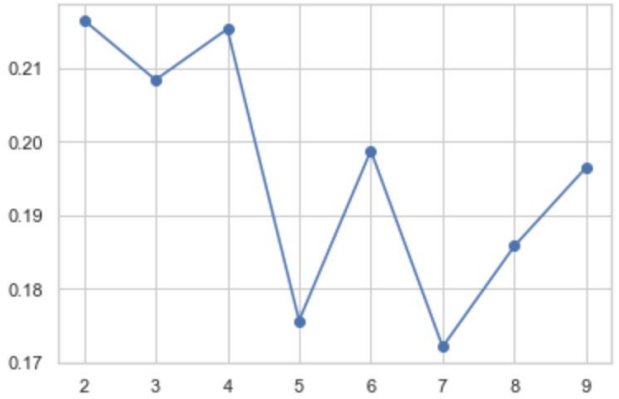


Caractérisation des clusters par les centroïdes K-Means

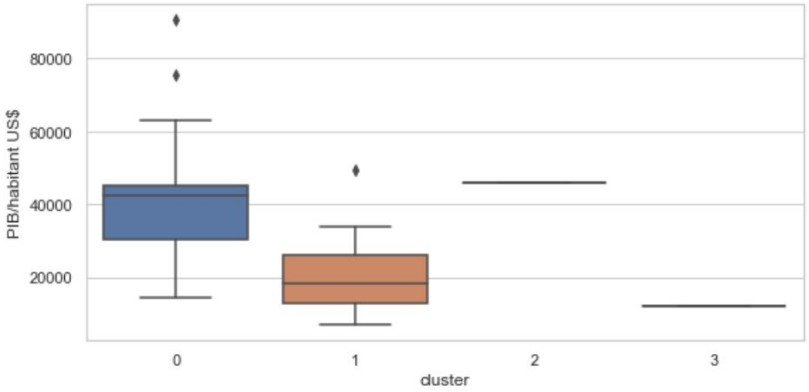
Nouveaux centroïdes sur le dernier K-Means avec k = 4. La métrique Silhouette expose un k optimal pour un découpage en 4 clusters.

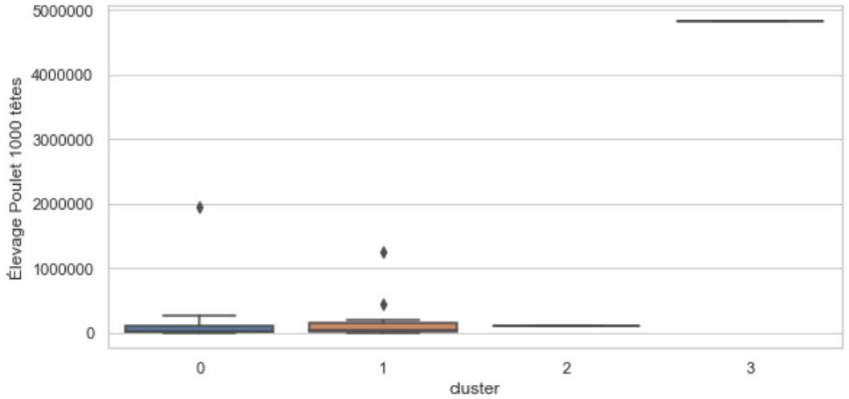
	PIB/habitant US$	Import Poulet en valeur 1000 US$	Élevage Poulet 1000 têtes	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	%Protéines animale/protéines totales	%Evolution Population 2012/2013
0	0.631236	-0.039870	-0.122685	0.690318	0.605072	0.283855	0.088477
1	-0.622194	-0.212084	-0.122946	-0.621628	-0.614106	-0.399038	-0.066974
2	0.842343	6.111468	-0.165614	-0.601235	0.799454	3.566951	-0.499058
3	-1.077451	0.439350	6.552021	-1.184713	-0.564567	-0.572189	-0.060021

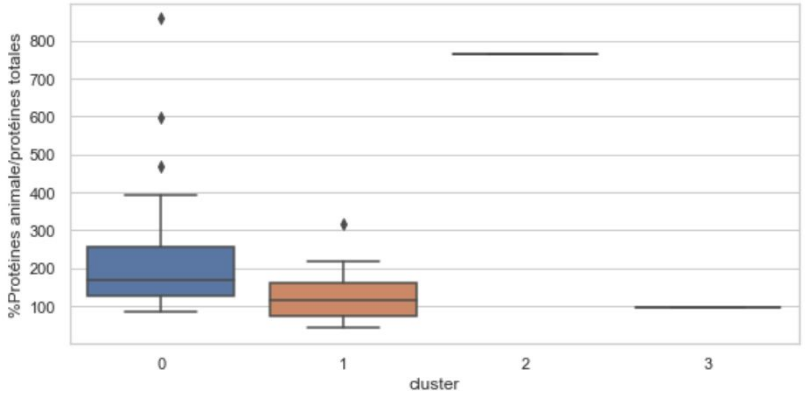
Le découpage a été réalisé en 4 clusters, la métrique Silhouette expose un bon équilibre dans ce clustering.

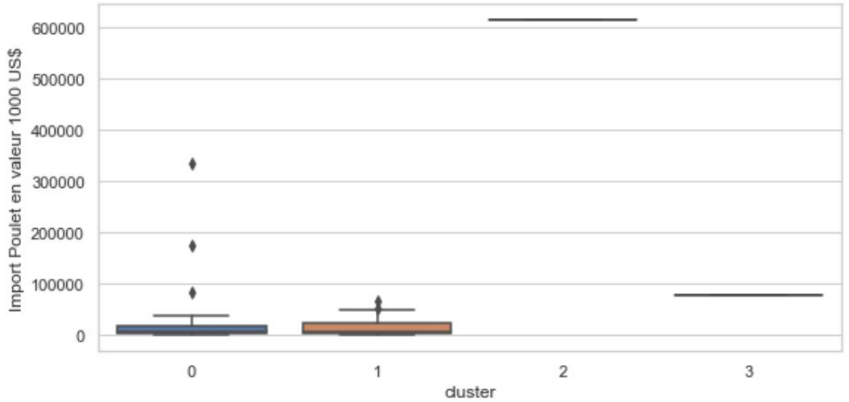


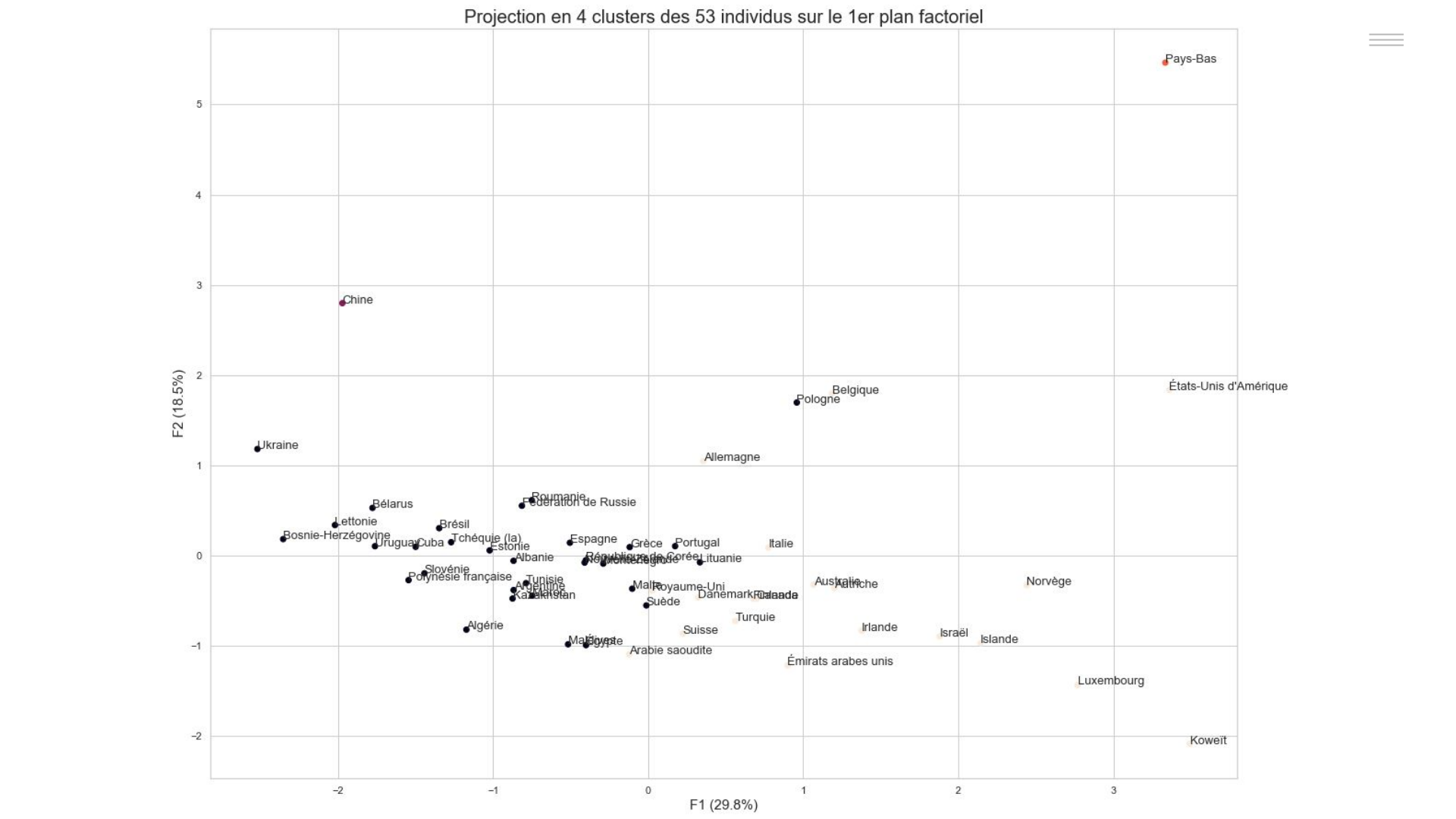
Boxplot de caractérisation des 4 clusters K-Means










Projection en 4 clusters des 53 individus sur le 1er plan factoriel
F1 (29.8%)
F2 (18.5%)
Pays-Bas
Chine
Belgique
Pologne
États-Unis d'Amérique
Ukraine
Allemagne
Roumanie
Bélarus
Fédération de Russie
Lettonie
Brésil
Bosnie-Herzégovine
Uruguay
Cuba
Tchéquie (la)
Estonie
Espagne
Grèce
Portugal
Italie
Albanie
Lituanie
Slovénie
Polynésie française
Tunisie
Malte
Royaume-Uni
Australie
Autriche
Norvège
Suède
Turquie
Algérie
Suisse
Irlande
Israël
Islande
Arabie saoudite
Émirats arabes unis
Luxembourg
Koweït

Interprétation et décision

L'étude de marché répond aux objectifs de l'entreprise. Ce cadre analytique se veut flexible, et reste ouvert à de nouvelles perspectives d'évolution stratégique.

- ➔ Des hypothèses d'aide à la décision sont proposées.
- ➔ Les premières réponses sont apportées.
- ➔ Des cibles géographiques seront proposées.

Hypothèses de réflexion sur les cibles possibles

Avant de pouvoir déterminer les pays cibles, il est indispensable de poser un cadre hypothétique. *L'identification du ou des clusters sera possible.*

«H0 : Les pays cibles ont un régime alimentaire riche en protéines animales.»

«H1 : Les pays cibles ont un régime alimentaire pauvre en protéines animales.»

«H2 : Les pays cibles enregistrent un PIB/habitant potentiellement prometteur.»

«H3 : Les pays cibles pratiquant l'import de poulets avec peu d'élevage.»

Cibles envisageables en réponse aux hypothèses

L'hypothèse nulle n'est pas rejetée pour les pays membres des clusters 2 et 3 ainsi que l'hypothèse H2 n'est pas rejetée pour le cluster 2 et 3 :

➔ Allemagne, Australie, Autriche, Belgique, Canada, Danemark, Finlande, Grèce, Irlande, Islande, Israël, Italie, Koweït, Lituanie, Luxembourg, Malte, Monténégro, Norvège, Pologne, Portugal, Royaume-Uni, Suisse, Suède, Turquie, Émirats arabes unis, États-Unis d'Amérique, Pays-Bas. **Pays avec un régime alimentaire riche en protéines animales et un PIB/habitant élevé.**

L'hypothèse nulle n'est pas rejetée en faveur de l'alternative H1 clusters 0 et 1 :

➔ Chine, Albanie, Algérie, Arabie saoudite, Argentine, Azerbaïdjan, Bosnie-Herzégovine, Brésil, Bélarus, Cuba, Espagne, Estonie, Fédération de Russie, Kazakhstan, Lettonie, Maldives, Maroc, Nouvelle-Zélande, Polynésie française, Roumanie, République de Corée, Slovénie, Tchéquie, Tunisie, Ukraine, Uruguay, Égypte. **Pays avec un régime alimentaire pauvre en protéines animales.**

L'hypothèse H3 n'est pas rejetée pour le cluster 2 :

➔ Pays-Bas. **Pays gros importateur de poulets, dont l'élevage reste limité.**

Contribution des pays dans l'inertie totale

Des pays sont plus importants que d'autres. Il est possible de quantifier le positionnement de chacun par l'analyse de leur inertie.

Les pays des **clusters 2 et 3** ont été retenus comme étant une cible attractive, **les plus représentatifs d'entre eux seront proposés.**

Pays-Bas	52.059815
Chine	46.336191
États-Unis d'Amérique	27.997383
Koweït	24.132815
Maldives	17.863746
Belgique	15.015129
Luxembourg	14.275649
Islande	10.523158
Israël	8.538410
Ukraine	8.366554
Pologne	7.988130
Norvège	7.756604
Lituanie	6.576089
Polynésie française	6.126091
Arabie saoudite	6.032434
Bosnie-Herzégovine	5.817174
Algérie	5.765709
Autriche	5.551419
Lettonie	5.476034
Turquie	4.801114
Égypte	4.620733
Allemagne	4.488161
Bélarus	4.474223
Émirats arabes unis	4.453005
Azerbaïdjan	4.371924
Uruguay	4.322590
Brésil	4.211997
Suisse	4.000994
Cuba	3.792647
Maroc	3.592358
Nouvelle-Zélande	3.496132
Irlande	3.263353
Tchéquie (la)	3.215479
Monténégro	3.166317
Australie	3.025658
Albanie	2.945046
Slovénie	2.655569
Finlande	2.623445

Pays cibles U.E

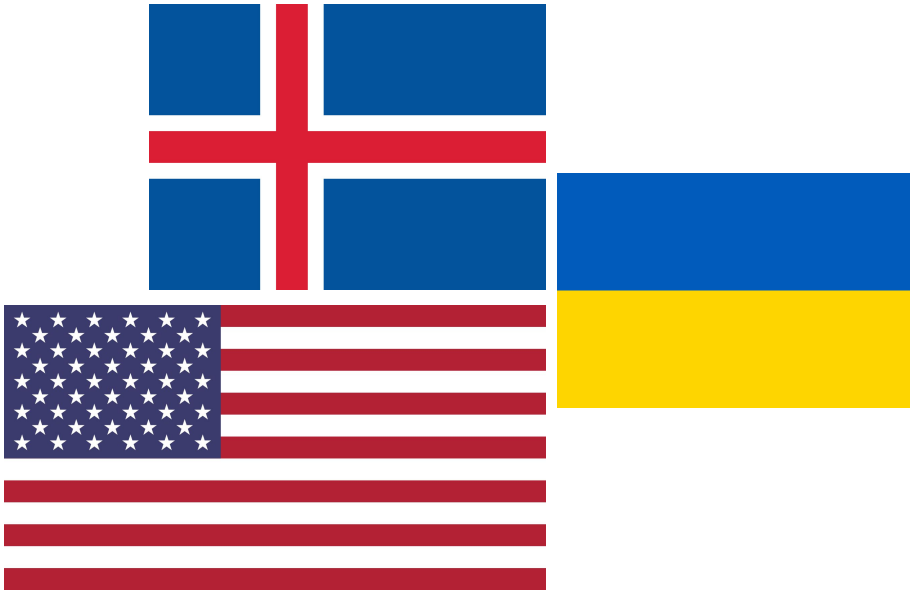
Cibles de l'Union Européenne (facilité monétaire, logistique, etc…), avec une position dominante en termes de dispo. en protéines animales, PIB/habitant, mais aussi sur la capacité d'importation de poulets vivants pour les Pays-Bas.

01 | Allemagne
02 | Belgique
03 | Luxembourg
04 | Pays-Bas
05 | Pologne

Pays	PIB/habitant US$	Import Poulet en valeur 1000 US$	Élevage Poulet 1000 têtes	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	%Protéines animale/protéines totales	%Evolution Population 2012/2013	cluster
Allemagne	42914.5	173124.0	160774.0	7001.0	202.98	103.48	-0.09	0
Belgique	41014.0	334425.0	36219.0	7470.0	198.97	86.85	0.40	0
Luxembourg	90656.4	850.0	111.0	7079.0	227.52	184.05	1.15	0
Pays-Bas	45753.6	614024.0	97719.0	6450.0	223.19	762.61	0.27	2
Pologne	23555.5	83107.0	117054.0	6901.0	202.93	598.93	0.02	0

Pays cibles hors U.E

Cibles possibles hors de l'Union Européenne, mais des contraintes d’échelle devront être attentivement mesurées, comme par exemple une barrière monétaire, logistique, etc...



01 | Islande
02 | Ukraine
03 | États-Unis d'Amérique

Pays	PIB/habitant US$	Import Poulet en valeur 1000 US$	Élevage Poulet 1000 têtes	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)	%Protéines animale/protéines totales	%Evolution Population 2012/2013	cluster
Islande	42372.0	0.0	217.0	6760.0	266.60	288.17	1.23	0
Ukraine	8338.9	64663.0	195256.0	6275.0	177.25	86.11	-0.64	1
États-Unis d'Amérique	51208.9	11616.0	1945900.0	7363.0	219.01	857.19	0.80	0

Test d'adéquation - Variable dont la loi est normale

Test d'adéquation de Kolmogorov-Smirnov :

On peut tester l'adéquation de la 'Disponibilité alimentaire (Kcal/personne/jour)' à une loi normale à l'aide de *Kolmogorov-Smirnov*. Le test sera doublé par celui de *Shapiro-Wilk*.

```
]: from scipy.stats import ks_2samp

   stat, p = ks_2samp(df_subset['Disponibilité alimentaire (Kcal/personne/jour)'],
             list(np.random.normal(np.mean(df_subset['Disponibilité alimentaire (Kcal/personne/jour)']),
             np.std(df_subset['Disponibilité alimentaire (Kcal/personne/jour)']), 1000)))
   print('Statistics=%.3f, p=%.3f' % (stat, p))

   #Interprétation
   alpha = 0.05
   if p > alpha:
       print('On ne peut pas rejeter H0 pour des niveaux de test de 5%')
   else:
       print('H0 est rejetée à un niveau de test de 5%')
```

```
Statistics=0.062, p=0.980
On ne peut pas rejeter H0 pour des niveaux de test de 5%
```

Test d'adéquation - Variable dont la loi est normale

Test d'adéquation de Shapiro-Wilk :

*Recommandé pour tester la normalité dans le cas de petits échantillons.

```
]: from scipy.stats import shapiro

   stat, p = shapiro(df_subset['Disponibilité alimentaire (Kcal/personne/jour)'])

   print('Statistics=%.3f, p=%.3f' % (stat, p))

   #Interprétation
   alpha = 0.05
   if p > alpha:
       print('On ne peut pas rejeter H0 pour des niveaux de test de 5%')
   else:
       print('H0 est rejetée pour des niveaux de test de 5%')
```

```
Statistics=0.981, p=0.529
On ne peut pas rejeter H0 pour des niveaux de test de 5%
```

-> Le Test de Shapiro-Wilk est plus précis que celui de Kolmogorov-Smirnov, et également plus adapté dans notre cas de petit échantillonnage. La variable 'Disponibilité de alimentaire exprimée en Kcal' suit une loi normale.

Test de comparaison - Clusters réellement distincts?

Test de comparaison de deux clusters dans le cas gaussien.

La variable 'Disponibilité alimentaire (Kcal/personne/jour)' suit une loi normale et sera par conséquent choisie pour le test.

]:
```
cluster_test1 = df_cls4[df_cls4['cluster'] == 1]['Disponibilité alimentaire (Kcal/personne/jour)']
cluster_test2 = df_cls4[df_cls4['cluster'] == 0]['Disponibilité alimentaire (Kcal/personne/jour)']
```

]:
```
#On teste tout d'abord l'égalité des variances à l'aide de la commande
from scipy.stats import bartlett
stat, p = bartlett(cluster_test1, cluster_test2)
print('Statistics=%.3f, p=%.3f' % (stat, p))

#Interprétation
alpha = 0.05
if p > alpha:
    print('On ne rejette donc pas H0, l'égalité des variances au niveau de test 5%')
else:
    print('H0 est rejetée au niveau de test 5%')
```

```
Statistics=0.186, p=0.667
On ne rejette donc pas H0, l'égalité des variances au niveau de test 5%
```

Test de comparaison - Clusters réellement distincts?

```
]: #On teste ensuite l'égalité des moyennes à l'aide de la commande
   from scipy.stats import ttest_ind
   stat, p = ttest_ind(cluster_test1, cluster_test2, equal_var=True)
   print('Statistics=%.3f, p=%.9f' % (stat, p))

   #Interprétation
   alpha = 0.05
   if p > alpha:
       print('On ne rejette donc pas H0, l'égalité des moyennes de nos 2 clusters au niveau de test 5%')
   else:
       print('H0 l\'hypothèse d'égalité des moyennes est rejetée au niveau de test 5%')
```

```
Statistics=-6.127, p=0.000000139
H0 l'hypothèse d'égalité des moyennes est rejetée au niveau de test 5%
```

-> On rejette que nos deux clusters suivent la même distribution, on a en effet rejeté l'hypothèse d'égalité des moyennes.

Les tests statistiques réalisés permettent de vérifier que nos clusters ne suivent pas la même distribution, en effet l'hypothèse d'égalité des moyennes a été rejetée dans le cadre du test précédent. Les clusters identifiés sont distincts.

Vision d’ici 2030

2020

Stratégie d’exportation

Travail préparatoire sur l’adaptation de l’offre et du marketing prix, packaging, nom…

2022

Exportation vers l’U.E francophone

Pays cibles francophones la Belgique et le Luxembourg, premières exportations.

2024

Exportation vers l’All. et la Pologne

Développement vers d’autres pays voisins avec une barrière linguistique à gérer en amont.

2026

Exportation vers les Pays-Bas

Conquête du marché Néerlandais en passant par Rotterdam le premier port d'Europe et l'un des plus important au monde.

2030

Exportation hors U.E

USA, Islande, Ukraine? De grands marchés à conquérir…