A large group of fluffy yellow chicks are gathered in a wooden brooder. The chicks are covered in soft, downy feathers and have small orange beaks. They are packed closely together, with some looking towards the camera and others looking away. The background is slightly blurred, emphasizing the chicks in the foreground.

Salomé Saffar, Data Analyst

# **Etude du marché du poulet à l'international**

# Le contexte

---

Entreprise française dans  
l'agroalimentaire

---

Stratégie de  
développement à  
l'international

---

Exportation de ses produits,  
pas de production dans les  
pays ciblés

# Problématique

Quels sont les pays propices à une insertion dans le marché du poulet ?

# Les données choisies



---

Croissance de la population entre 2016 et 2017

---

---

PIB par habitant

---

---

Disponibilité alimentaire en calories  
(Kcal/personne/jour)

---

---

Disponibilité de protéines en quantité  
(g/personne/jour)

---

---

Proportion de protéines d'origine animale  
dans la disponibilité de protéines total

---

---

Proportion de protéines issue de la viande  
de volailles

---

# Méthodes d'analyse utilisées

Classification hiérarchique:  
dendrogramme

Méthode de classification : k-means

ACP : analyse en composante principale

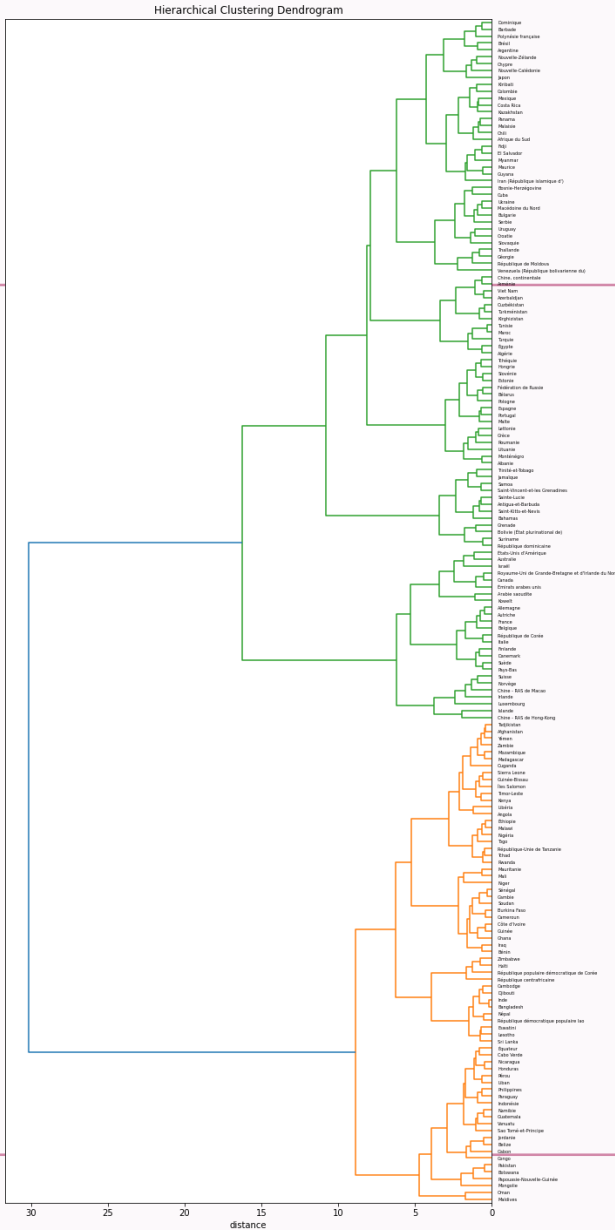
Tests statistiques :

- Test d'adéquation : Kolmogorov Smirnov
- Test de comparaison sur deux clusters choisis : égalité des variances et des moyennes

# Création des clusters

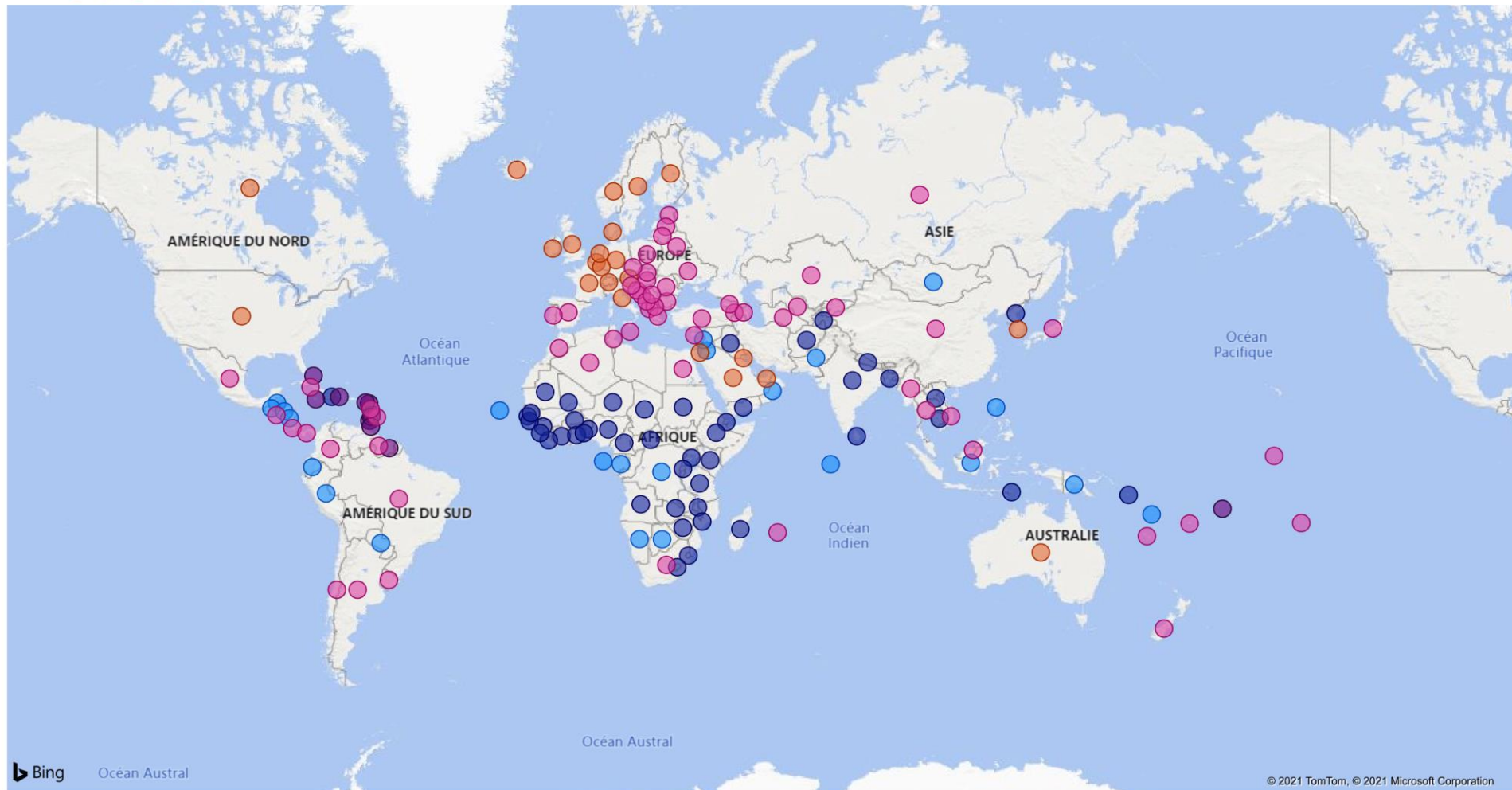
# Le dendrogramme

- Création de 5 clusters



pays et cluster

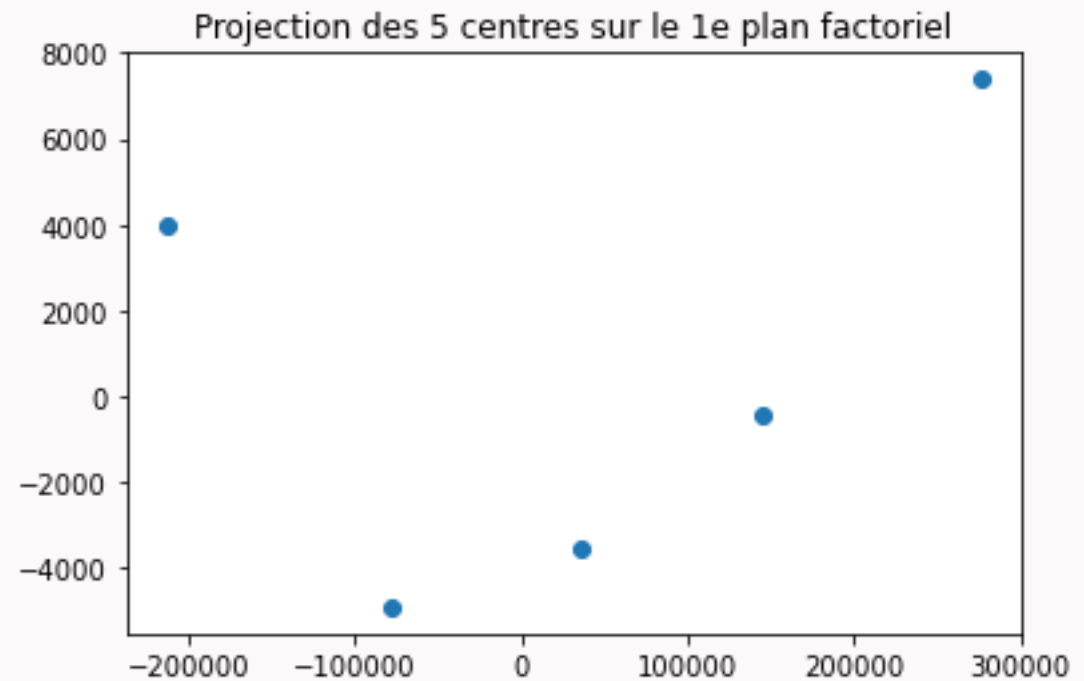
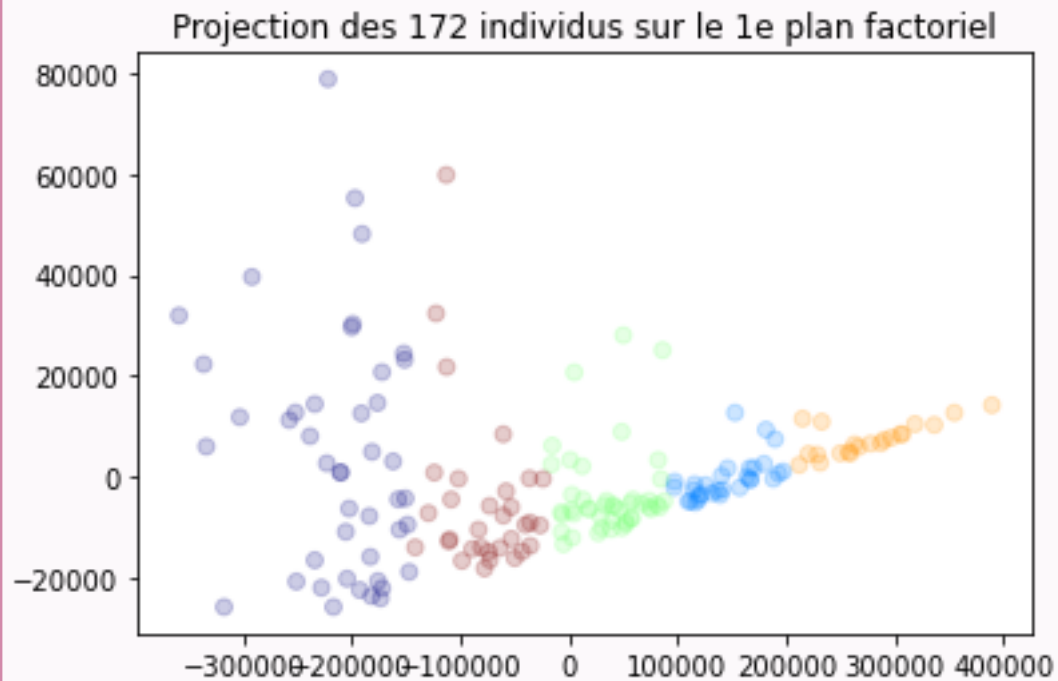
cluster 1 2 3 4 5





# Algorithme K-Means

Identification de deux clusters qui se démarquent



# Les centres de classes

	pays	croissance_pop	PIB_hab	prot_anim_prot_tot	prot_hab	prot_volaille	kcal_hab	cluster
0	Arménie	0.29	3914.525246	46.83	35634.95	5.34	1118360.0	5
1	Afghanistan	2.58	513.085978	19.88	20494.75	1.00	747155.0	2
2	Albanie	-0.08	4514.204908	54.37	43106.50	3.76	1236985.0	5
3	Algérie	2.07	4109.701336	28.92	33788.05	2.44	1216910.0	5

Permet de caractériser chacun des clusters et d'en choisir deux à étudier plus en profondeur

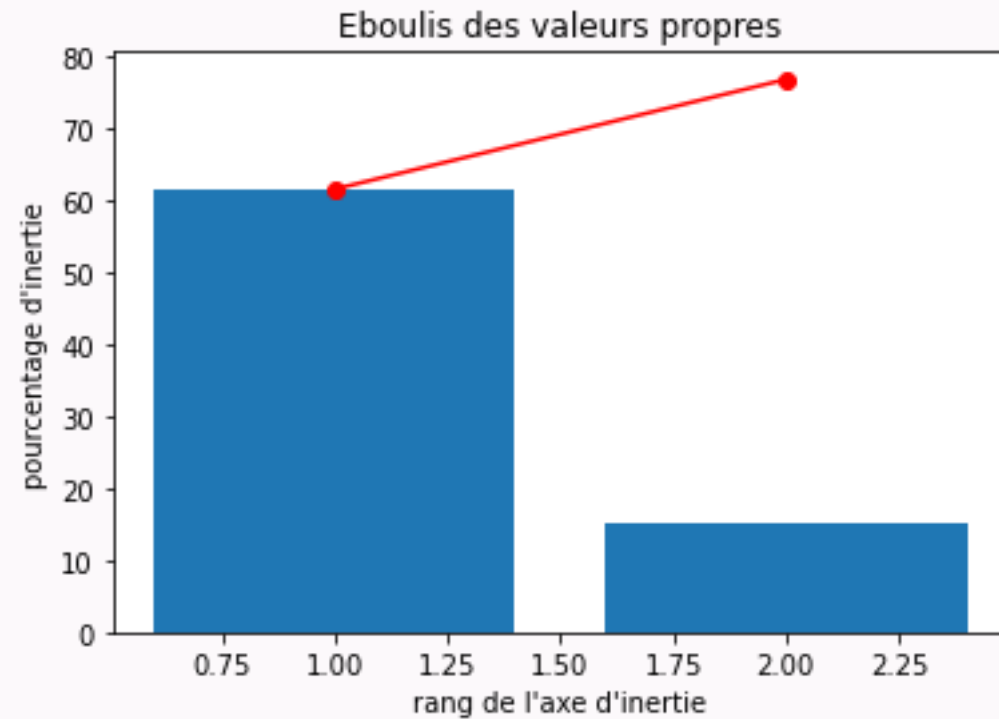
# Caractéristiques des clusters

- Cluster 1 : 23 pays
- Cluster 2 : 47 pays
- Cluster 3 : 25 pays
- Cluster 4 : 12 pays
- Cluster 5 : 65 pays

Nous allons sélectionner en priorité les pays du **cluster 3** et dans un second temps les pays du **cluster 5**.

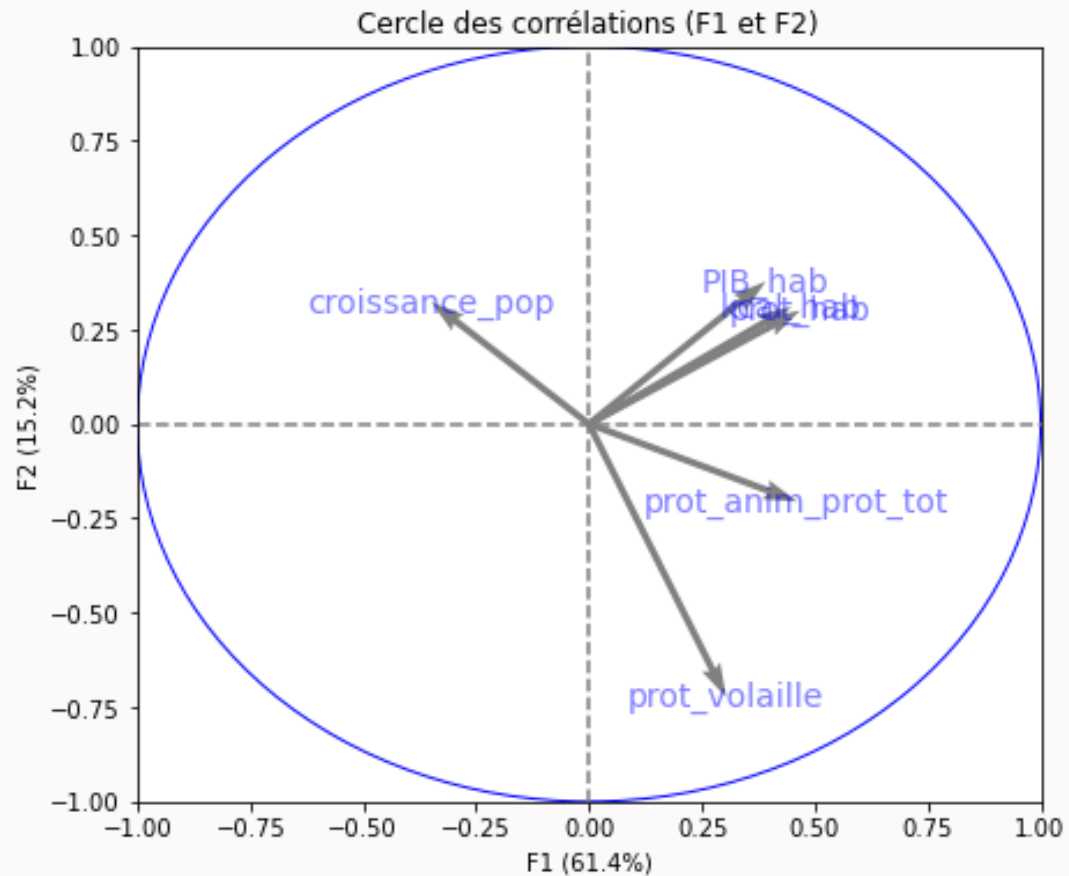
# Analyse en composante principale

- Critère de Kaiser =  $(100/6) = 16.6\%$
- Conservation des composantes F1 et F2
- Besoin de deux composantes minimum afin de constituer un plan



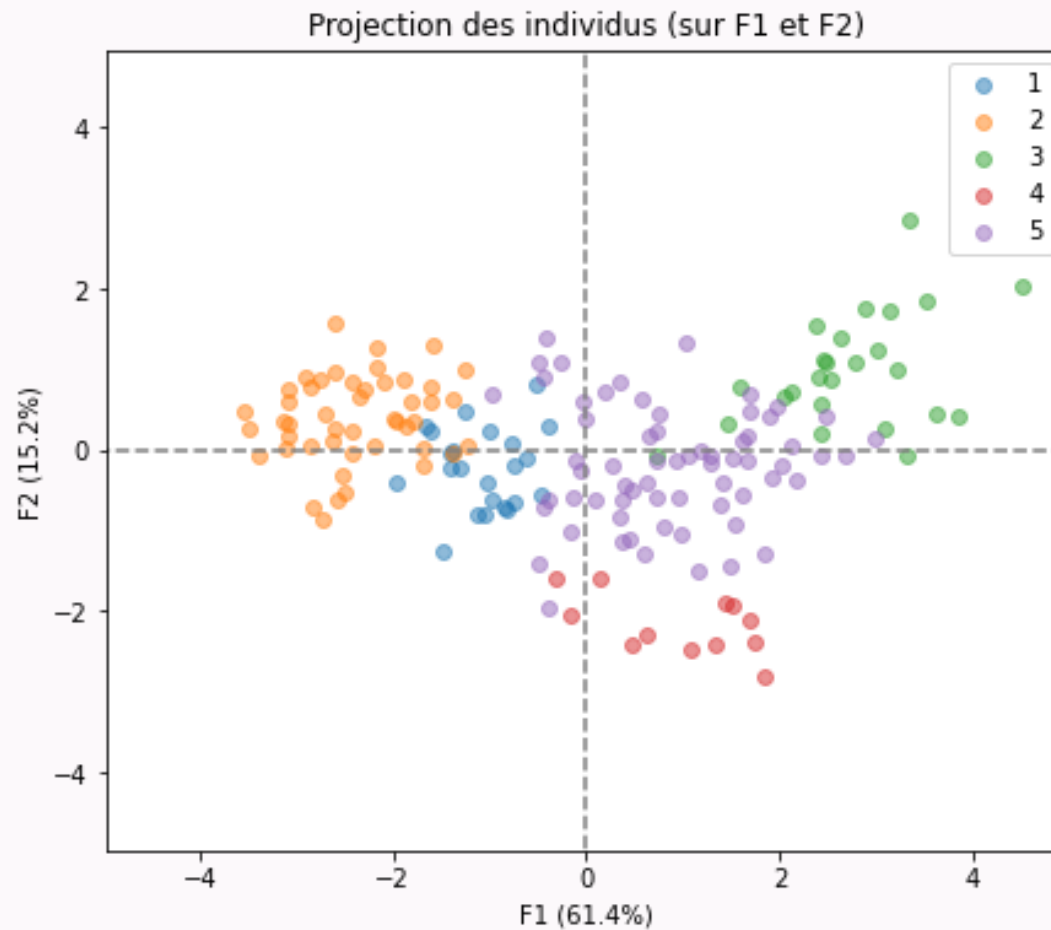
# Cercle des corrélations

- La croissance de la population contribue fortement à la formation de F2
- F1 réunit les variables concernant la disponibilité alimentaire
- Corrélation entre le PIB par habitant et la disponibilité alimentaire en protéines et en calories



# Projection des individus

- Cluster 3 qui se démarque
- Intérêt pour le cluster 5 qui se confirme



# Tests statistiques

# Kolmogorov Smirnov

- Variables continues
- $\alpha = 0.05$
- $N = 172$
- $1.358 \div \sqrt{172} = 0.10$

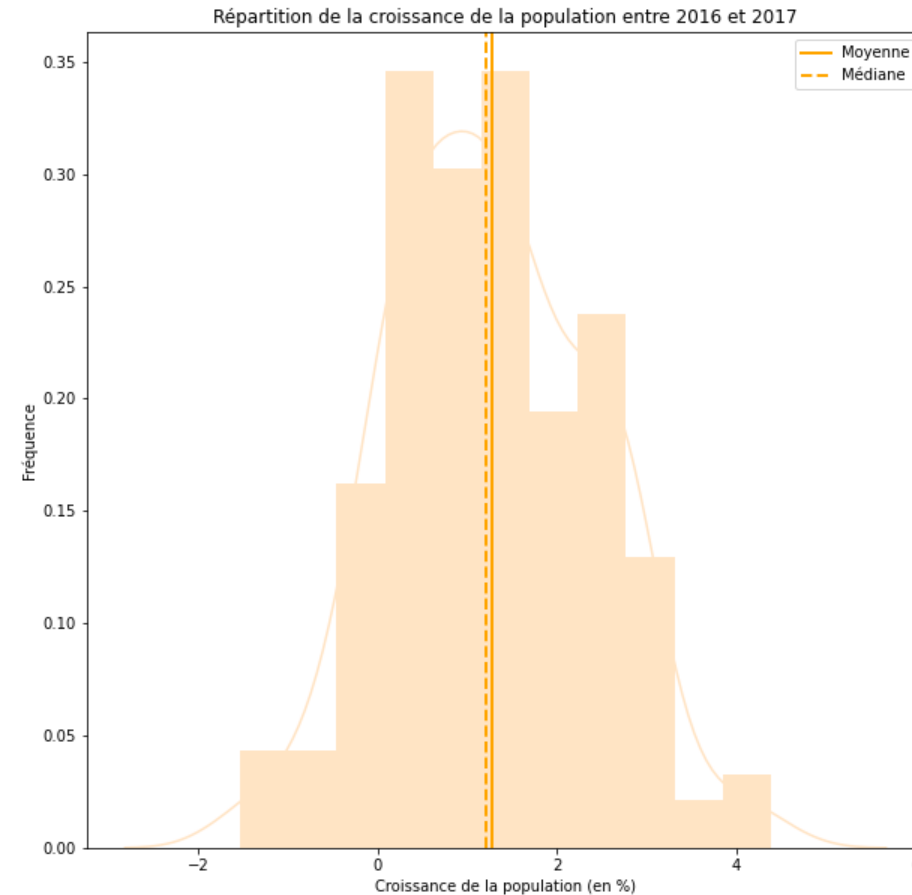
Notre valeur D doit être inférieur à 0.10 afin de déterminer que la distribution suit une loi normale (hypothèse H0)

	$\alpha$				
n	0.001	0.01	0.02	0.05	0.1
1		0.99500	0.99000	0.97500	0.95000
2	0.97764	0.92930	0.90000	0.84189	0.77639
3	0.92063	0.82900	0.78456	0.70760	0.63604
4	0.85046	0.73421	0.68887	0.62394	0.56522
5	0.78137	0.66855	0.62718	0.56327	0.50945
6	0.72479	0.61660	0.57741	0.51926	0.46799
7	0.67930	0.57580	0.53844	0.48343	0.43607
8	0.64098	0.54180	0.50654	0.45427	0.40962
9	0.60846	0.51330	0.47960	0.43001	0.38746
10	0.58042	0.48895	0.45662	0.40962	0.36866
11	0.55588	0.46770	0.43670	0.39122	0.35242
12	0.53422	0.44905	0.41918	0.37543	0.33815
13	0.51490	0.43246	0.40362	0.36143	0.32548
14	0.49753	0.41760	0.38970	0.34890	0.31417
15	0.48182	0.40420	0.37713	0.33760	0.30397
16	0.46750	0.39200	0.36571	0.32733	0.29471
17	0.45440	0.38085	0.35528	0.31796	0.28627
18	0.44234	0.37063	0.34569	0.30936	0.27851
19	0.43119	0.36116	0.33685	0.30142	0.27135
20	0.42085	0.35240	0.32866	0.29407	0.26473
25	0.37843	0.31656	0.30349	0.26404	0.23767
30	0.34672	0.28988	0.27704	0.24170	0.21756
35	0.32187	0.26898	0.25649	0.22424	0.20184
40	0.30169	0.25188	0.23993	0.21017	0.18939
45	0.28482	0.23780	0.22621	0.19842	0.17881
50	0.27051	0.22585	0.21460	0.18845	0.16982
Over 50	$1.94947/\sqrt{n}$	$1.62762/\sqrt{n}$	$1.51743/\sqrt{n}$	$1.35810/\sqrt{n}$	$1.22385/\sqrt{n}$



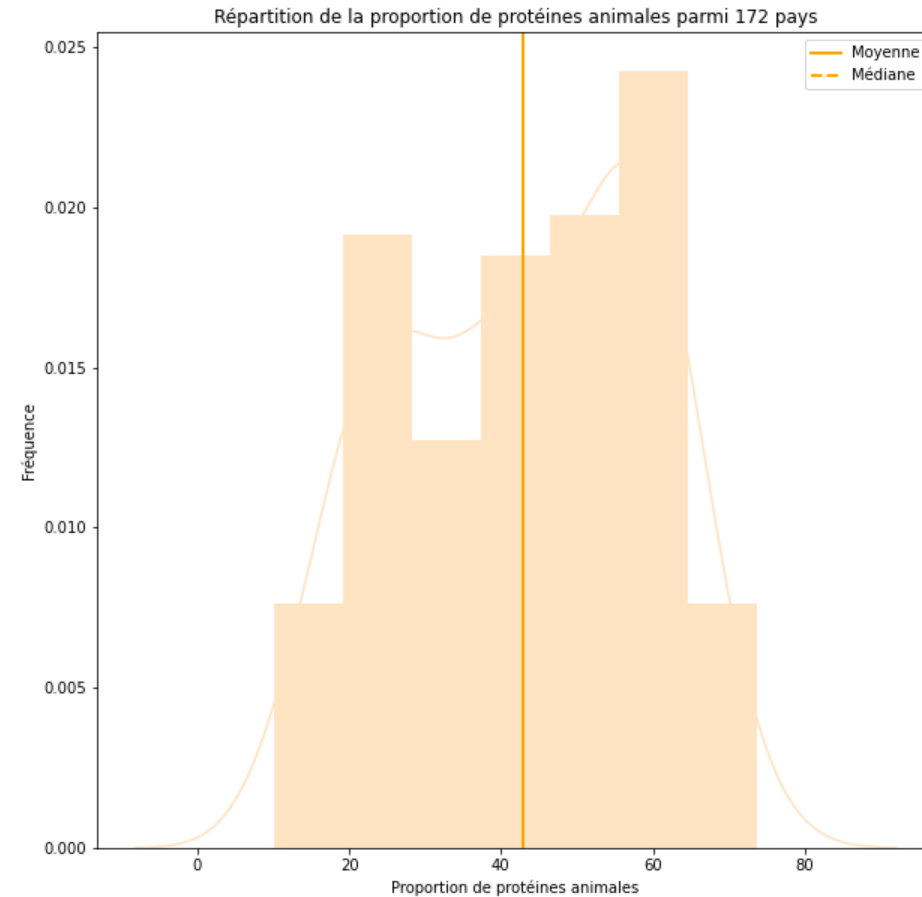
# Test d'adéquation : la croissance de la population

- Kolmogorov Smirnov
- $D = 0.05$
- On ne peut pas rejeter l'hypothèse nulle  $H_0$
- La croissance de la population suit la loi normale



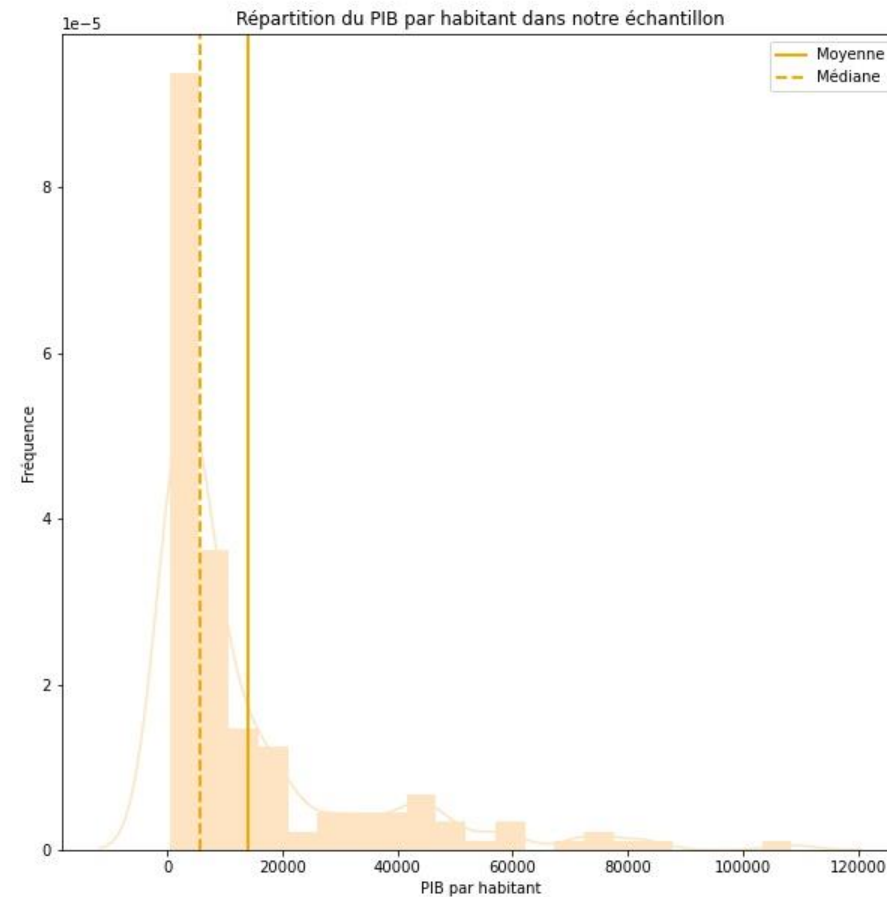
# Test d'adéquation : la proportion de protéines animales

- Kolmogorov Smirnov
- $D = 0.088$
- On ne peut pas rejeter l'hypothèse nulle  $H_0$
- La proportion de protéines animales suit la loi normale



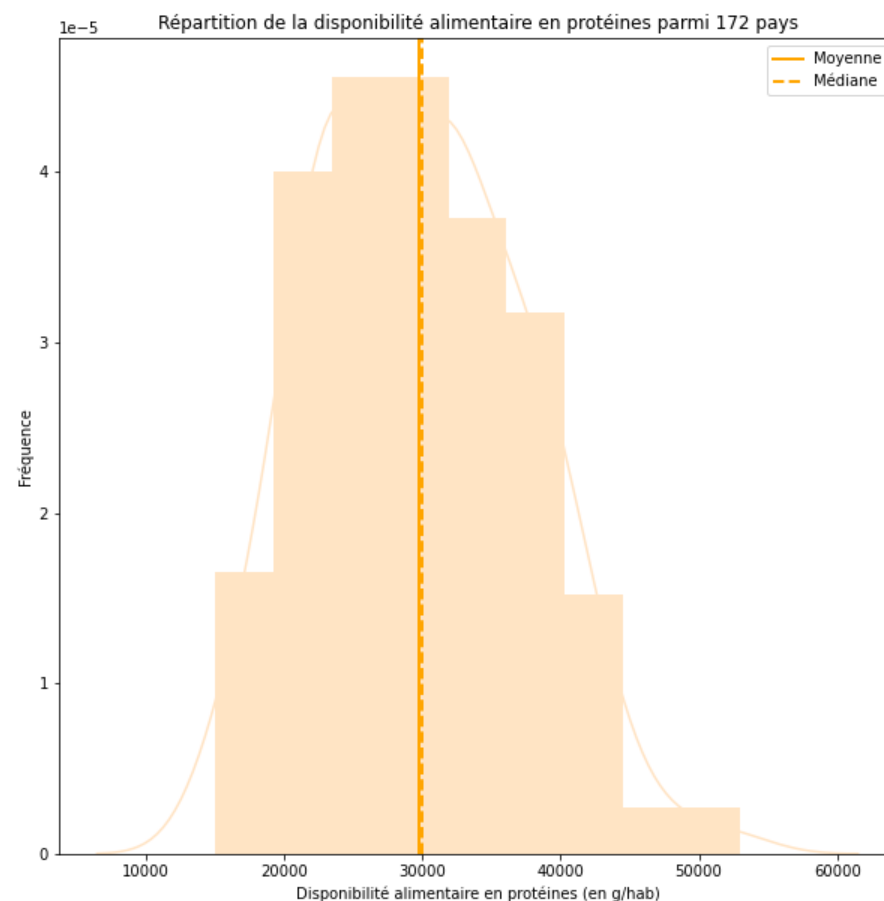
# Test d'adéquation : le PIB par habitant

- Kolmogorov Smirnov
- $D = 0.244$
- On ne peut rejeter l'hypothèse nulle  $H_0$
- Le PIB par habitant ne suit pas la loi normale



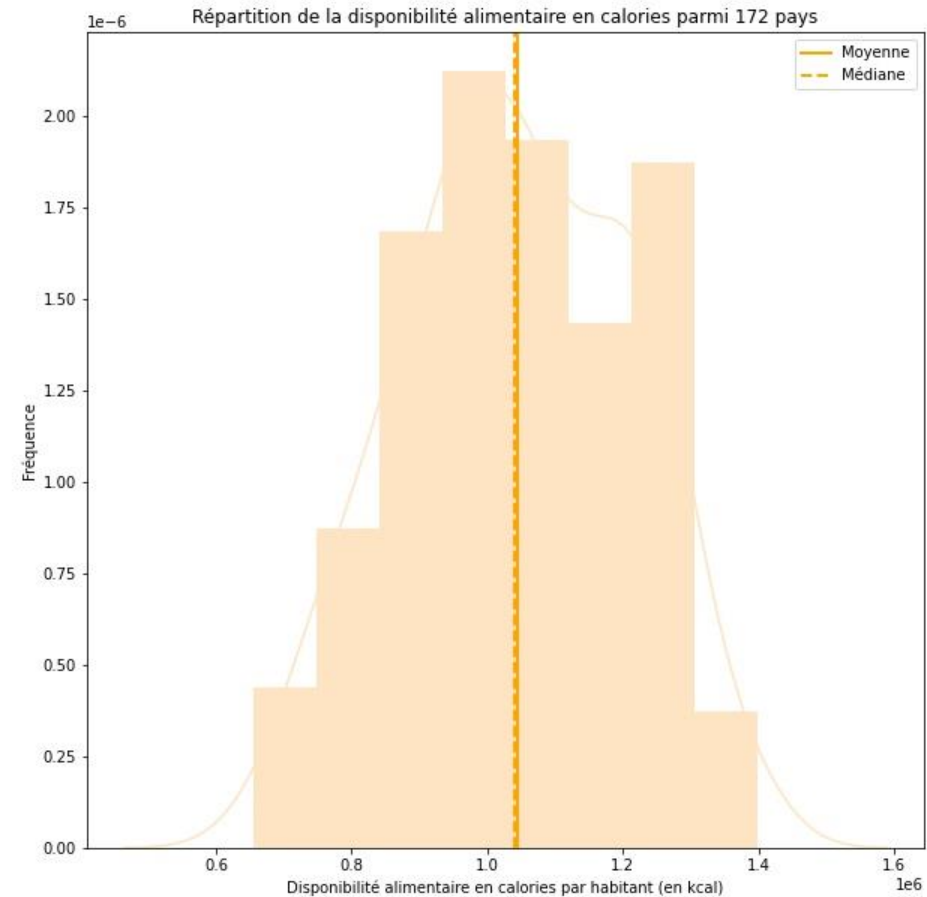
# Test d'adéquation : la disponibilité alimentaire en protéines

- Kolmogorov Smirnov
- $D = 0.06$
- On ne peut pas rejeter l'hypothèse nulle  $H_0$
- La disponibilité alimentaire en protéines suit la loi normale



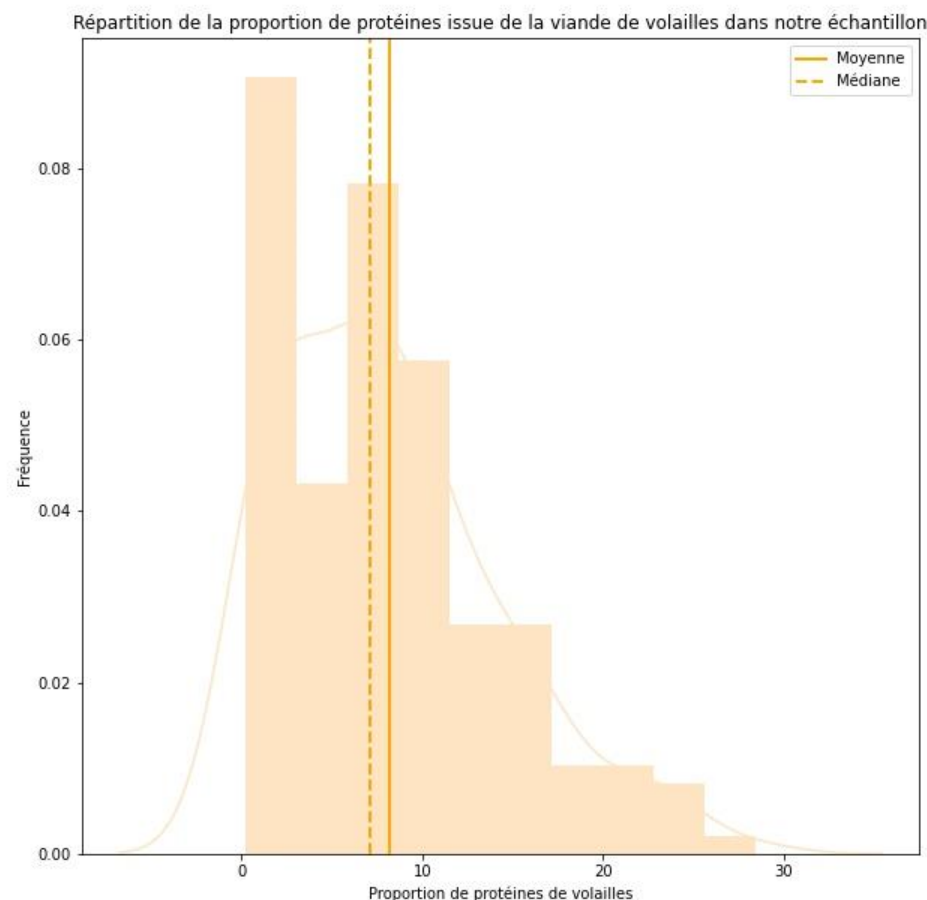
# Test d'adéquation : la disponibilité alimentaire en calories

- Kolmogorov Smirnov
- $D = 0.07$
- On ne peut pas rejeter l'hypothèse nulle  $H_0$
- La disponibilité alimentaire en calories suit la loi normale



# Test d'adéquation : la proportion de protéines issue de la viande de volailles

- Kolmogorov Smirnov
- $D = 0.09$
- On ne peut pas rejeter l'hypothèse nulle  $H_0$
- La proportion de protéines issue de la viande de volailles suit la loi normale



# Test de comparaison : égalités des variances et des moyennes

- Choix des clusters 3 et 5
- Egalité des variances : Test de Barlett si la distribution suit la loi normale
- Egalité des moyennes : Test de Student en cas d'égalité des variances sinon test de Welch\*
- Dans le cas où la distribution d'une variable ne suit pas une loi normale, on effectuera le test de Wilcoxon-Mann-Whitney

*\*Le test t de Welch est une adaptation du test t de Student. Il peut être utilisé notamment pour tester statistiquement l'hypothèse d'égalité de deux moyennes avec deux échantillons de variances inégales. Il s'agit en fait d'une solution approchée du problème de Behrens–Fisher.*

# Déterminer si la différence est statistiquement significative

Valeur de  $p \leq \alpha$  : la différence entre les moyennes ou des variances est statistiquement significative (Rejeter  $H_0$ )

Valeur de  $p > \alpha$  : la différence entre les moyennes ou des variances n'est pas statistiquement significative (Impossible de rejeter  $H_0$ )

$$\alpha = 0,05$$



# Test de comparaison : la croissance de la population

## EGALITÉ DES VARIANCES

P-value = 0,12

On ne peut pas rejeter l'hypothèse d'égalité des variances : Test de Student

## EGALITÉ DES MOYENNES : TEST DE STUDENT

P-value = 0,016

On peut rejeter l'hypothèse d'égalité des moyennes

Différence significative des moyennes de la croissance de la population entre le cluster 3 et le cluster 5

# Test de comparaison : la proportion de protéines animales

## EGALITÉ DES VARIANCES

P-value = 0,22

On ne peut pas rejeter l'hypothèse d'égalité des variances : Test de Student

## EGALITÉ DES MOYENNES : TEST DE STUDENT

P-value =  $7,038 \times 10^{-6}$

On peut rejeter l'hypothèse d'égalité des moyennes

La moyenne de la proportion de protéines animales entre les deux clusters est différente

# Test de comparaison : le PIB par habitant

Test de Mann-Whitney

P-value =  $5.53 \times 10^{-13}$

On ne peut pas rejeter l'hypothèse d'égalité des variances

# Test de comparaison : la disponibilité alimentaire en protéines

## EGALITÉ DES VARIANCES

P-value = 0,5

On ne peut pas rejeter l'hypothèse d'égalité des variances

## EGALITÉ DES MOYENNES : TEST DE STUDENT

P-value =  $8,49 \times 10^{-8}$

On peut rejeter l'hypothèse d'égalité des moyennes

La moyenne de la disponibilité alimentaire en protéines entre les deux clusters est différente

# Test de comparaison : la disponibilité alimentaire en calories

## EGALITÉ DES VARIANCES

P-value = 0,016

On peut rejeter l'hypothèse d'égalité des variances

## TEST DE WELCH

P-value =  $3,98 \times 10^{-8}$

On peut rejeter l'hypothèse d'égalité  $H_0$

La moyenne de la disponibilité alimentaire en calories entre les deux clusters est différente

# Test de comparaison : la proportion de protéines de volaille

## EGALITÉ DES VARIANCES

P-value = 0,56

On ne peut pas rejeter l'hypothèse d'égalité des variances

## EGALITÉ DES MOYENNES : TEST DE STUDENT

P-value = 0,21

On ne peut pas rejeter l'hypothèse d'égalité des moyennes

La moyenne de la proportion de protéines issue de la viande de volaille entre les deux clusters est identique

# Conclusion et recommandations

# Les pays sélectionnés



*Islande*



*Norvège*



*USA*



*Irlande*



*Lituanie*



*Israël*



*Danemark*



*Australie*



*Suisse*



*Luxembourg*



# Avantages

- Pays ouvert à l'importation
- Privilégier les pays européens de la liste : logistique facilitée, libre-échange au sein de l'Union Européenne
- Pays à haut potentiel commerciale
- Situation politique stable
- USA et Australie : opportunité économique développée et une clientèle potentiellement très élargie
- Valorisation du Made in France dans ces pays
- Etude des pays européens et du Moyen-Orient du cluster 5 pouvant être également des opportunités économiques : Chypre, Estonie, Koweït, Malte, Pologne, Portugal, Arabie Saoudite...