

# POULETS FRAIS D'EVRY

Etude de marché à l'export

# DONNÉES

**FAO**

- Démographiques
- Alimentaires



**nettoyage**

**analyse descriptive**

**Classification**

**ACP**



**Python**

**Exportation**

# NETTOYAGE :

- Vérification valeurs manquantes
- Imputation par la moyenne (10% partie alimentaire)

# DONNÉES

country	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	code	evolution_pop	protein_ani_ratio
Afghanistan	2675.545301	72.282003	AFG	12.5	0.146094
Afrique du Sud	3258.418120	91.019791	ZAF	6.2	0.379698
Albanie	3874.981711	132.904634	ALB	-0.7	0.499004
Algérie	3616.418120	100.509791	DZA	8.5	0.265223
Allemagne	3830.418120	111.729791	DEU	1.8	0.564576

171 obs

Démographie

Nutrition

Protéines

- Zone
- Code zone

- evolution\_pop  
(13-17)

- disponibilité alimentaire

- disponibilité
- Ratio animal

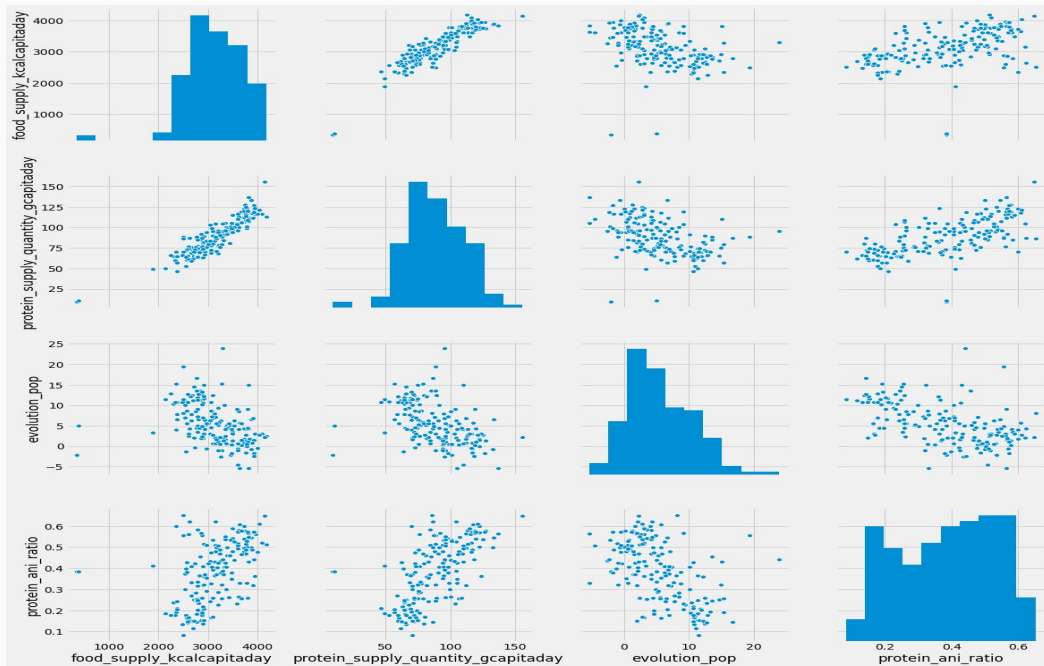
# PARTIE 1

# **ANALYSE DESCRIPTIVE**

# TENDANCES CENTRALES

	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio
count	171.000000	171.000000	171.000000	171.000000
mean	3128.054157	88.277969	5.452632	0.383632
std	575.017569	22.526994	5.046050	0.145564
min	339.272651	9.574739	-5.400000	0.083064
25%	2781.918120	71.446002	1.850000	0.256227
50%	3127.199916	86.949895	4.900000	0.397125
75%	3600.736283	104.767265	9.100000	0.507515
max	4177.127181	155.702108	23.900000	0.651428

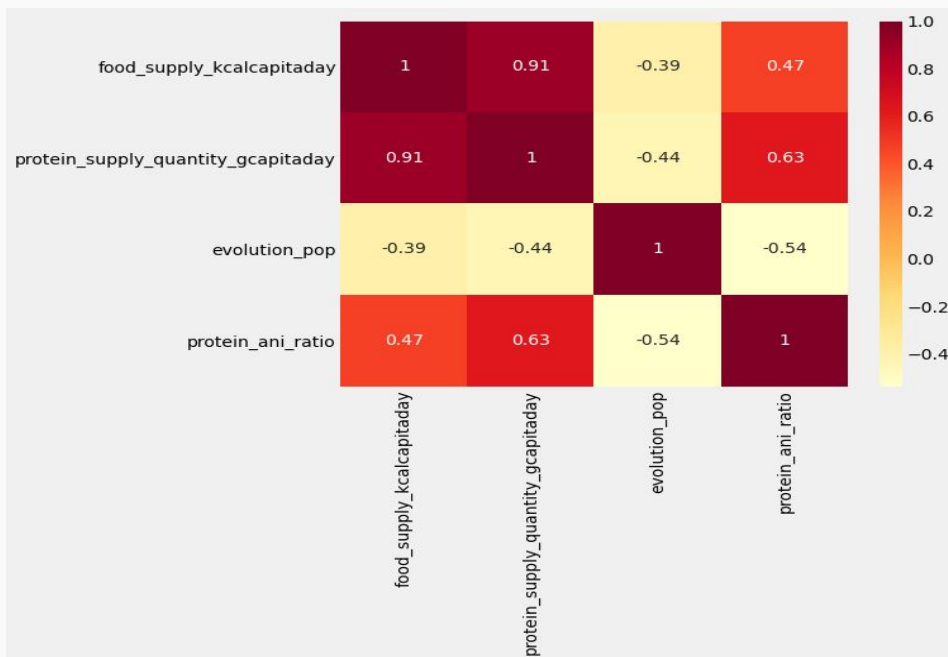
# MATRICE PAR PAIRES



- Corrélations
- Distributions asymétriques
- Comparaisons



# MATRICE CORRELATIONS

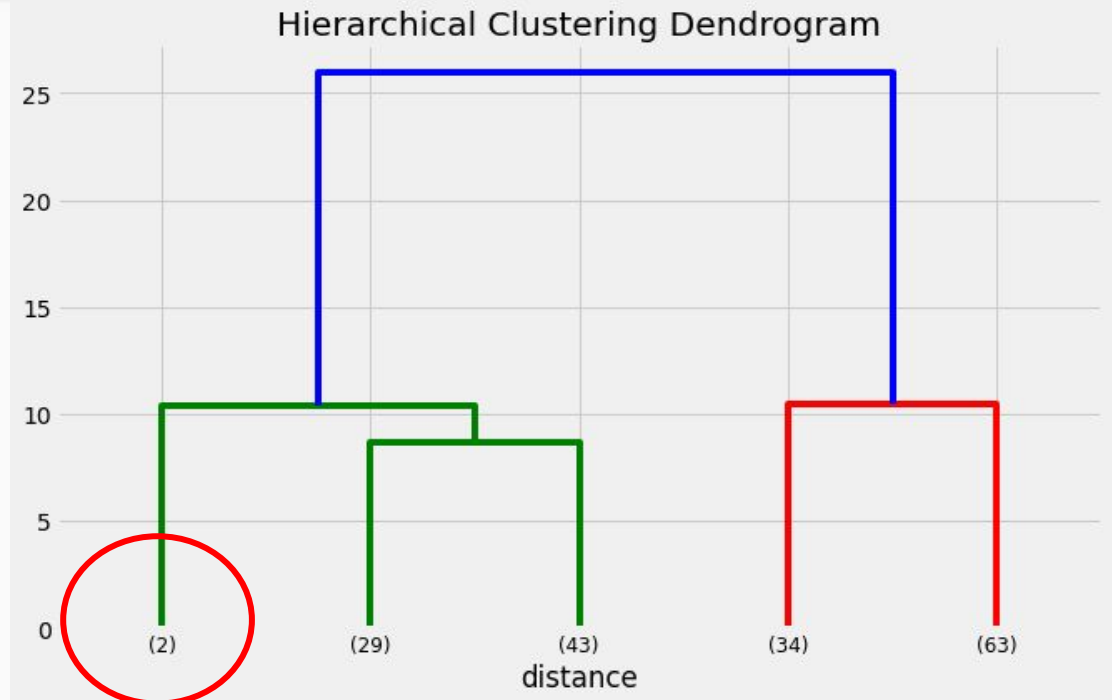


- Corrélations
- Classification

# **CLASSIFICATION HIÉRARCHIQUE**

# DENDROGRAMME

Cluster n = 2 ?



# Cluster 1

## Nb individus

	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio
clusters				
1	2	2	2	2
2	29	29	29	29
3	43	43	43	43
4	34	34	34	34
5	63	63	63	63

## Moyenne clusters

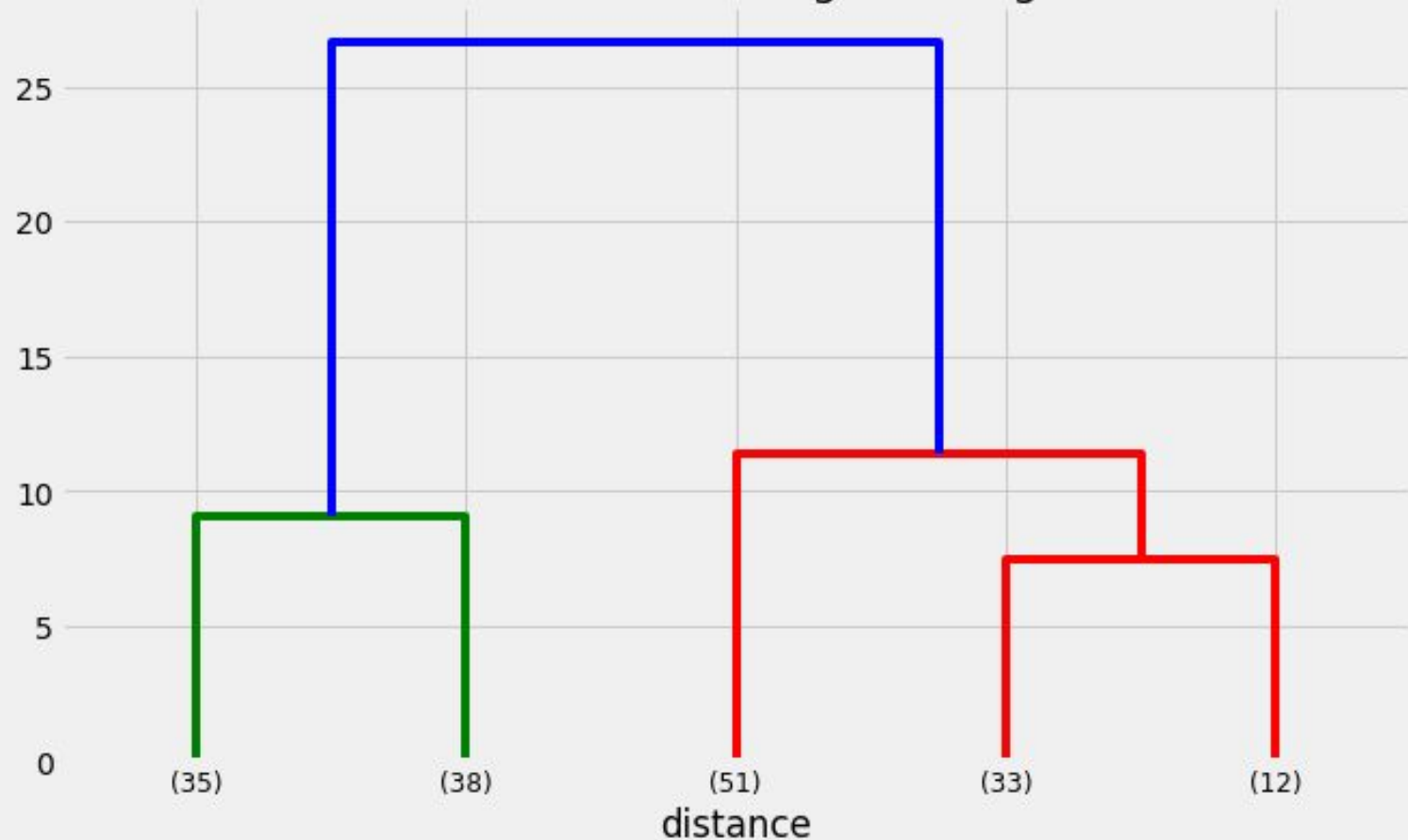
	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio
clusters				
1	356.236283	10.053475	1.450000	0.383632
2	2773.489594	72.487039	5.237931	0.309910
3	2719.147083	69.301600	11.920930	0.221379
4	3791.447507	117.468531	0.800000	0.538025
5	3300.334131	95.228616	3.774603	0.444989

- Bermudes
- Brunéi Darussalam

Valeurs aberrantes

-> suppression

# Hierarchical Clustering Dendrogram

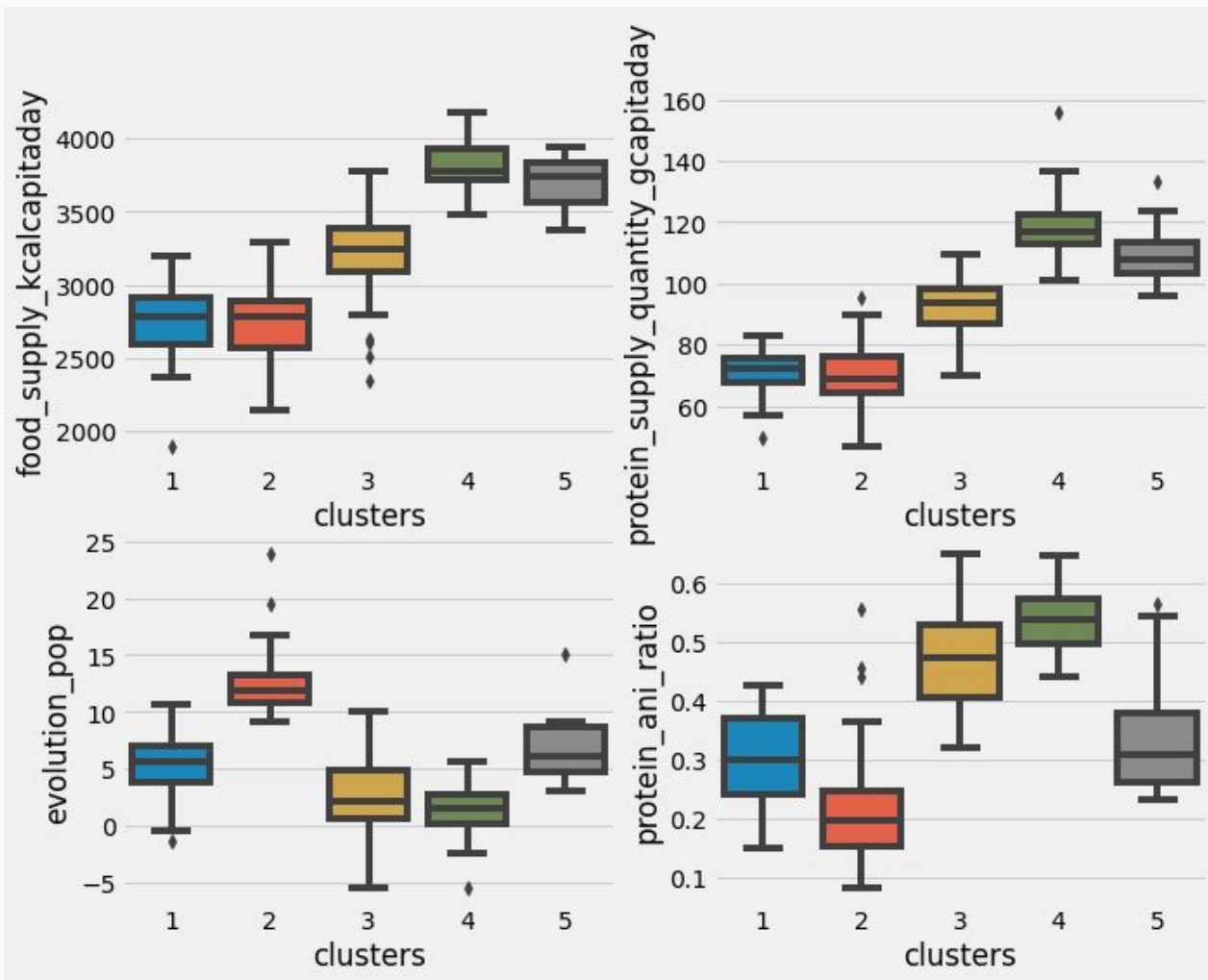


# DESCRIPTION CLUSTERS

	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio
clusters				
1	2751.395783	71.265442	5.488571	0.299120
2	2738.633908	70.192428	12.576316	0.222640
3	3208.750737	91.995562	2.613725	0.470676
4	3813.216994	118.274633	1.318182	0.537877
5	3694.616582	109.915546	6.891667	0.345829

## Cluster 4

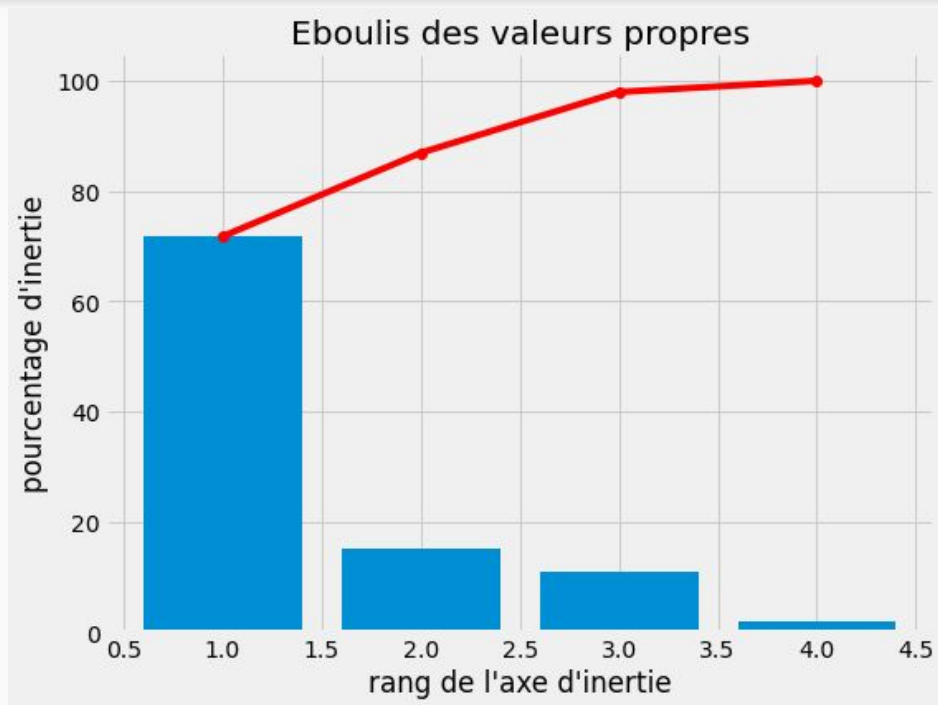
France (33 pays)



**ACP**

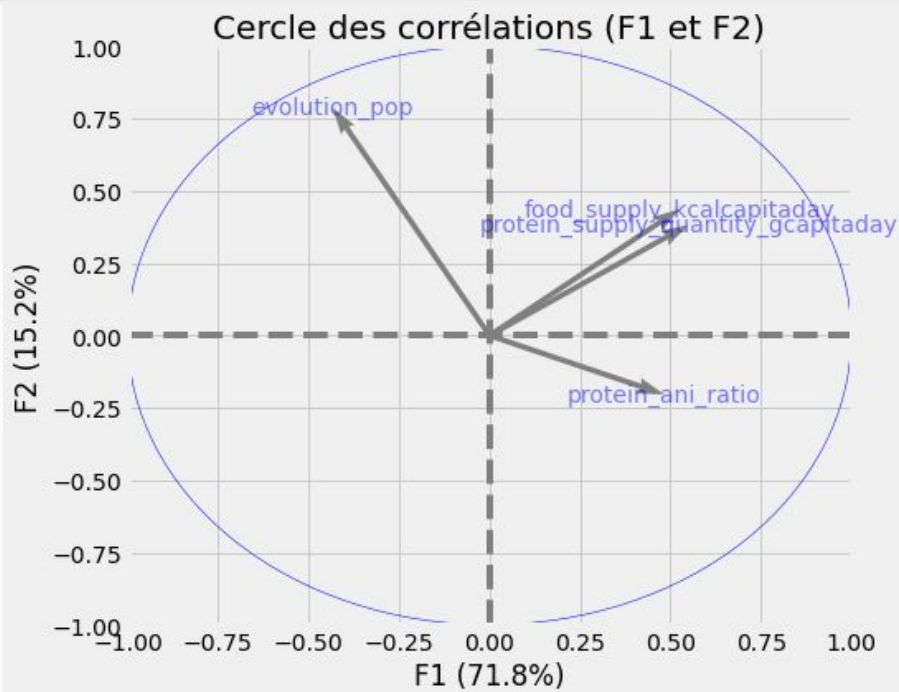


# EBOULIS VALEURS PROPRES



Choix de conserver composante 1  
(méthode coude - critère de Kaiser,  
 $100/p$ , soit 50%)

# CERCLE CORRELATIONS



**Disponibilité**

0,52

**Protéines**

0,54

**Ratio**

0,48

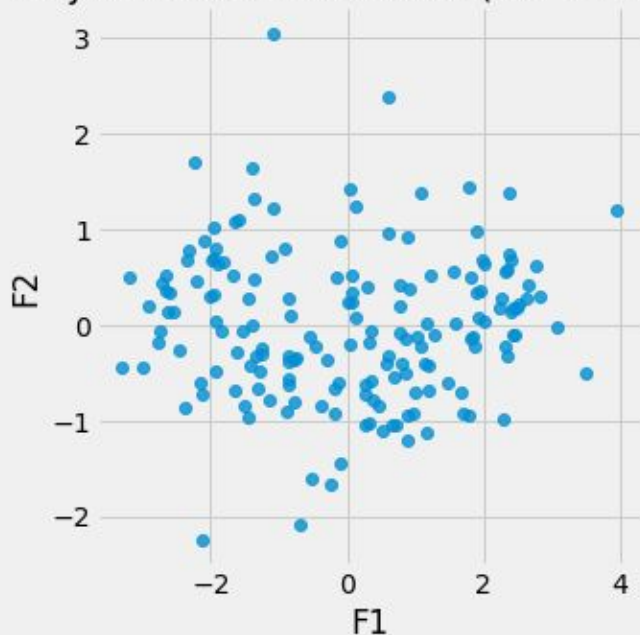
**Population**

-0,43

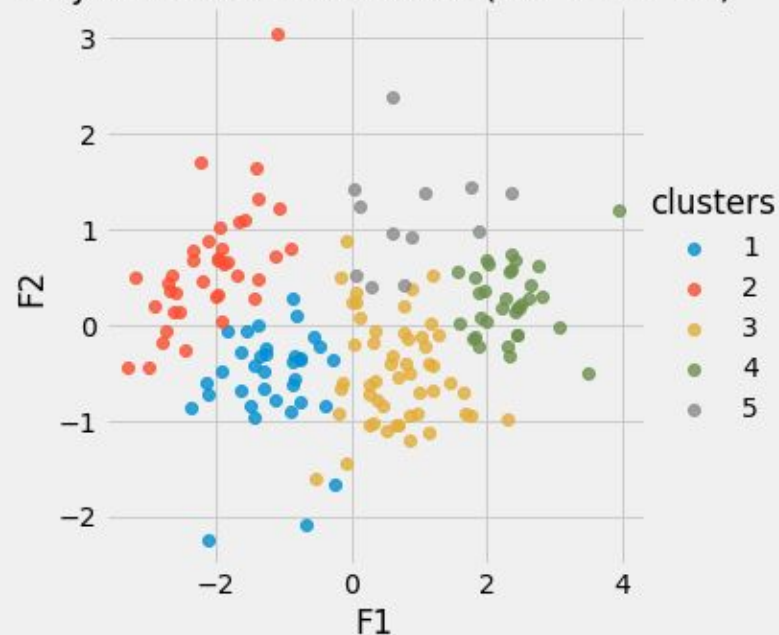
F1 : Relation données alimentation  
et augmentation population

# PROJECTION DES INDIVIDUS ET CLUSTERS

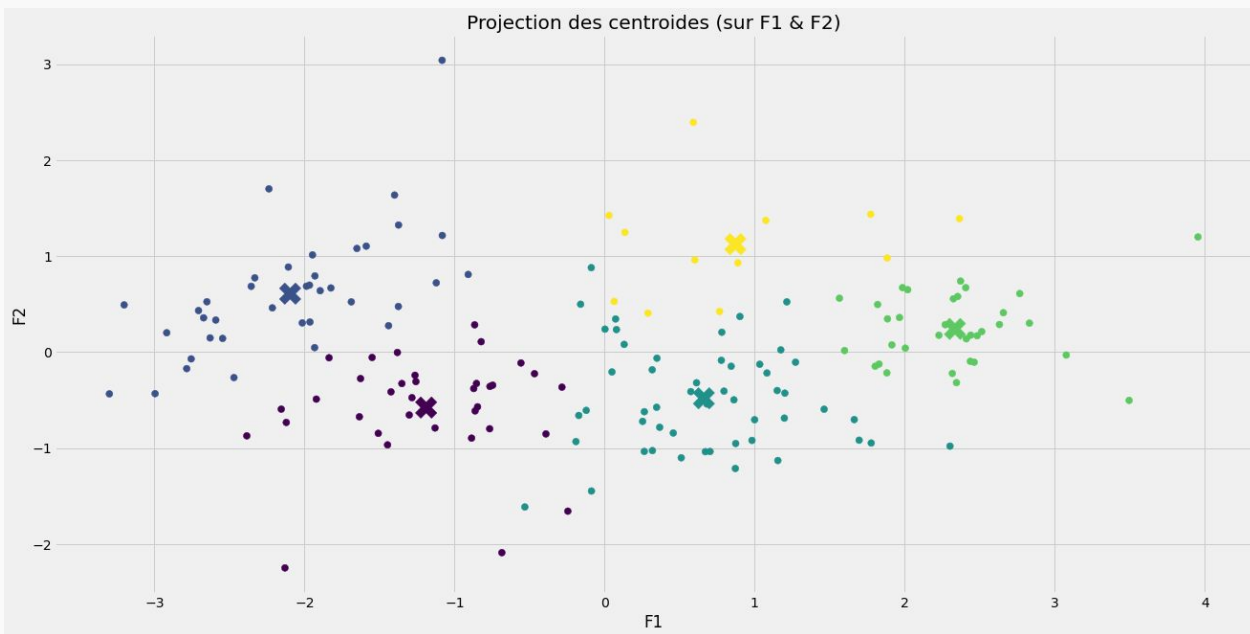
Projection des individus (sur F1 & F2)



Projection des individus (sur F1 & F2)

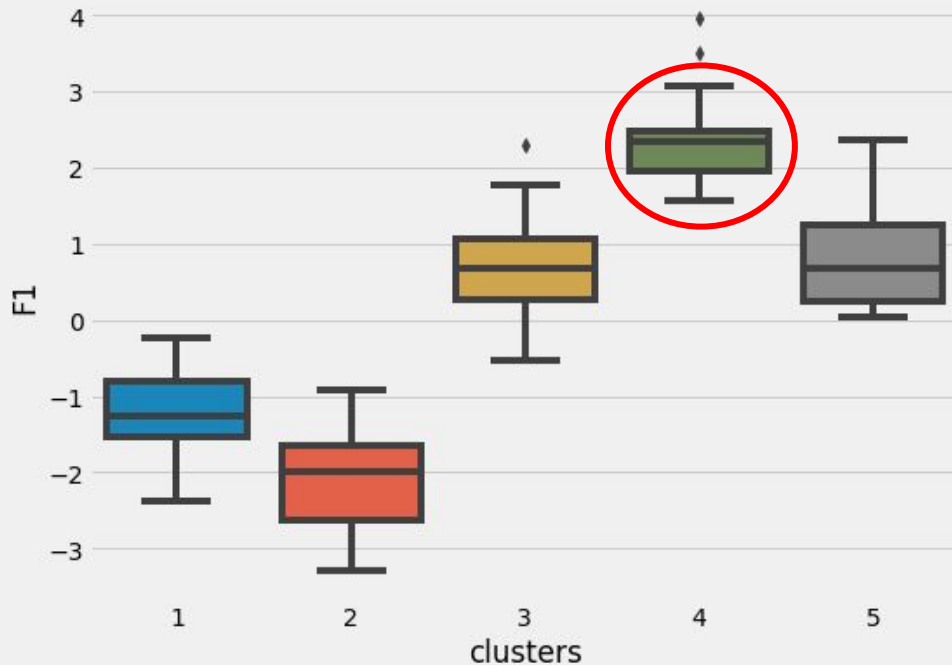


# PROJECTION CENTROÏDES



	F1	F2
clusters		
1	-1.189067	-0.577031
2	-2.095513	0.609748
3	0.661652	-0.480089
4	2.334608	0.242551
5	0.871709	1.125504

# CARACTÉRISATION CLUSTERS

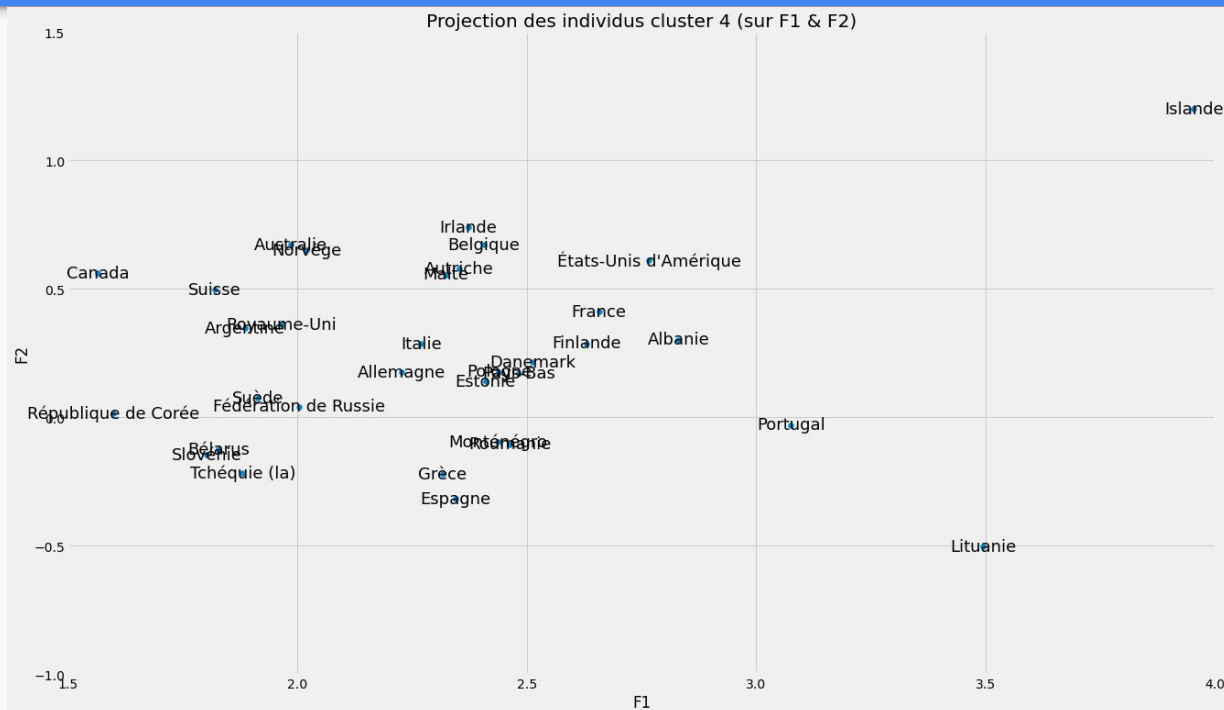


Cluster 4 semble le plus intéressant

Réserves:

- Comparaison variable
- Nombre individus (33)

# LISTE PAYS



- Hors UE
- Hors Europe
- Politiques commerciales

# COMPARAISONS VARIABLES

## **Adhésion loi normale :**

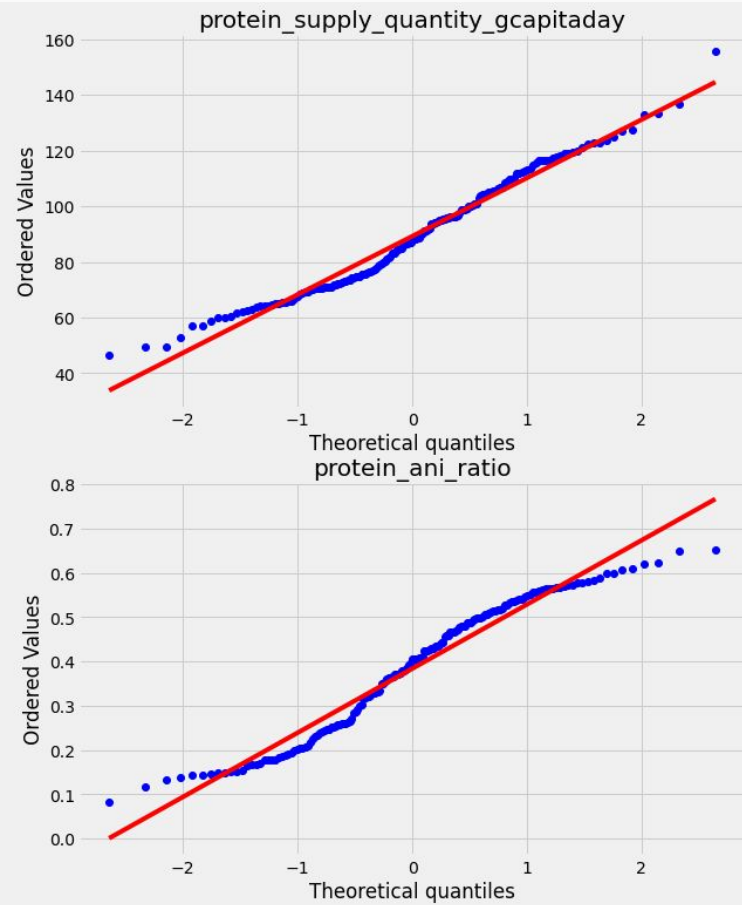
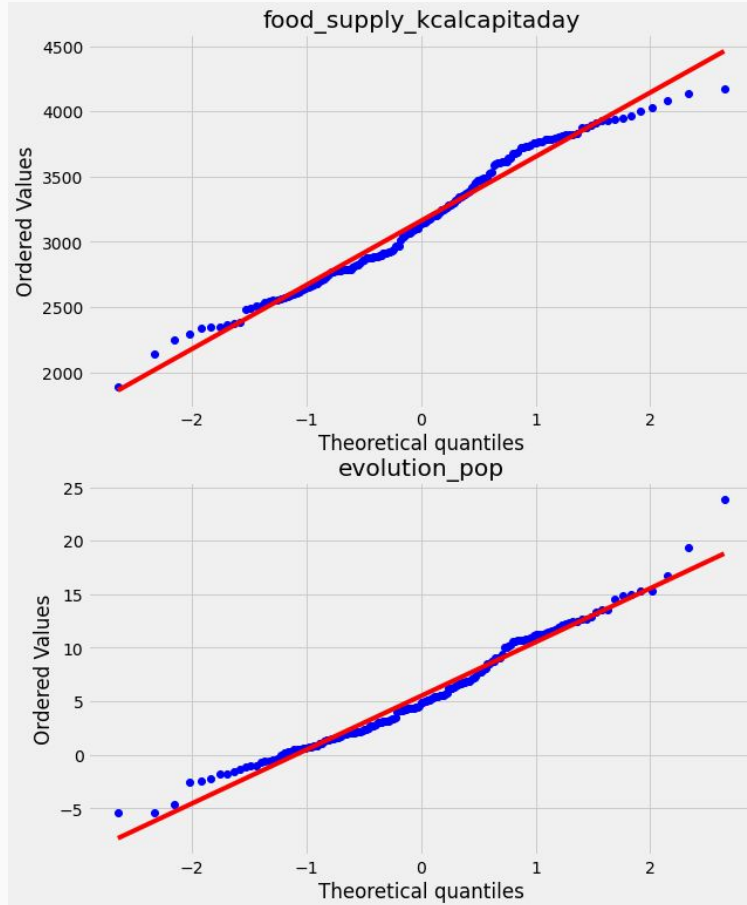
- (Histogramme)
- Shapiro-Wilk
- Droite de Henry

## **Egalité de variance :**

- Bartlett

## **Comparaison moyenne :**

- t-test





# Adhésion loi normale

## Shapiro-wilk

H0 : adhésion loi normale

H1 : pas adhésion loi normale

Seuil : 5%

**Disponibilité**

$p = 0,006$

**Population**

$p = 0,004$

**Protéines**

$p = 0,005$

**Ratio**

$p = 2,65e-5$

**On peut rejeter l'hypothèse de normalité au niveau de test 5% pour l'ensemble des variables**

# **PARTIE 2**

## **(Ajouts nouvelles variables)**

# DONNÉES

## FAO

- Démographiques
- Alimentaires

## WorldBank

- Revenus médian



**nettoyage**

**analyse descriptive**

**Classification**

**ACP**

# NETTOYAGE :

## Revenus

- Vérification valeurs manquantes
- Suppression (Paradis fiscaux, micro-états, instabilité)

# DONNÉES

country	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	code	evolution_pop	protein_animal_ratio	revenu	import_kgcapitaday
Islande	4137.908976	155.702108	ISL	2.2	0.648032	59619.513720	0.016386
Portugal	3935.054446	126.857160	PRT	-1.8	0.573086	17188.420795	0.021836
Albanie	3874.981711	132.904634	ALB	-0.7	0.499004	4105.150515	0.036097
États-Unis d'Amérique	4033.418120	121.269791	USA	2.7	0.609220	51484.941243	0.001037
France	3929.199916	122.612212	FRA	1.5	0.568133	32671.755086	0.021380

31 obs

Démographie

Nutrition

Protéines

Economique

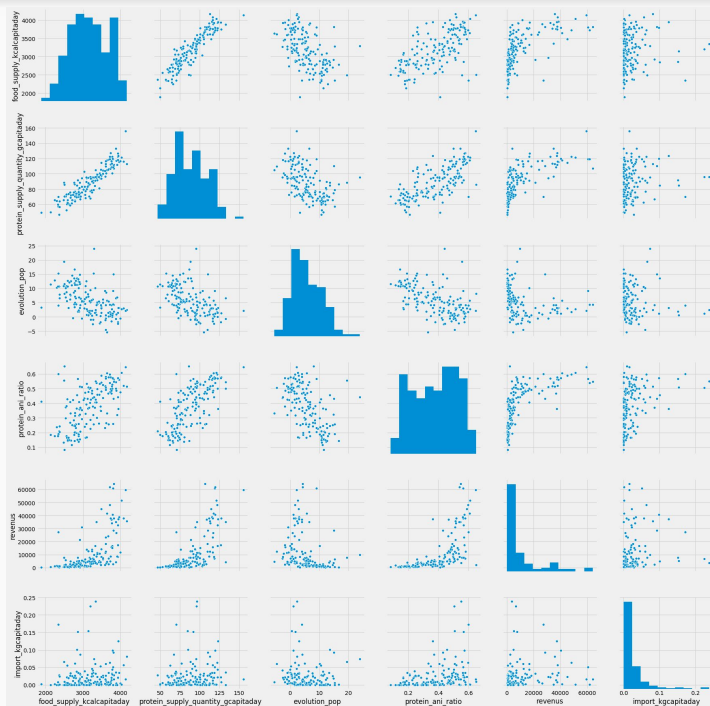
- Zone
- Code zone
- evolution\_pop (13-17)
- disponibilité alimentaire
- disponibilité
- Ratio animal
- Import
- Revenus médian

# **ANALYSE DESCRIPTIVE**

# TENDANCES CENTRALES

	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio	revenus	import_kgcapitaday
count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000
mean	3168.041655	88.948372	5.728846	0.377594	10531.436593	0.023813
std	495.087683	20.921784	5.070904	0.144604	14554.306691	0.038438
min	1889.709060	46.612317	-5.400000	0.083064	242.937497	0.000000
25%	2787.972672	71.885396	1.875000	0.251979	1364.108249	0.001125
50%	3146.454488	88.184791	5.100000	0.387206	4148.906447	0.009083
75%	3607.663548	105.171555	9.575000	0.499902	12359.080737	0.028245
max	4177.127181	155.702108	23.900000	0.651428	64306.962196	0.238418

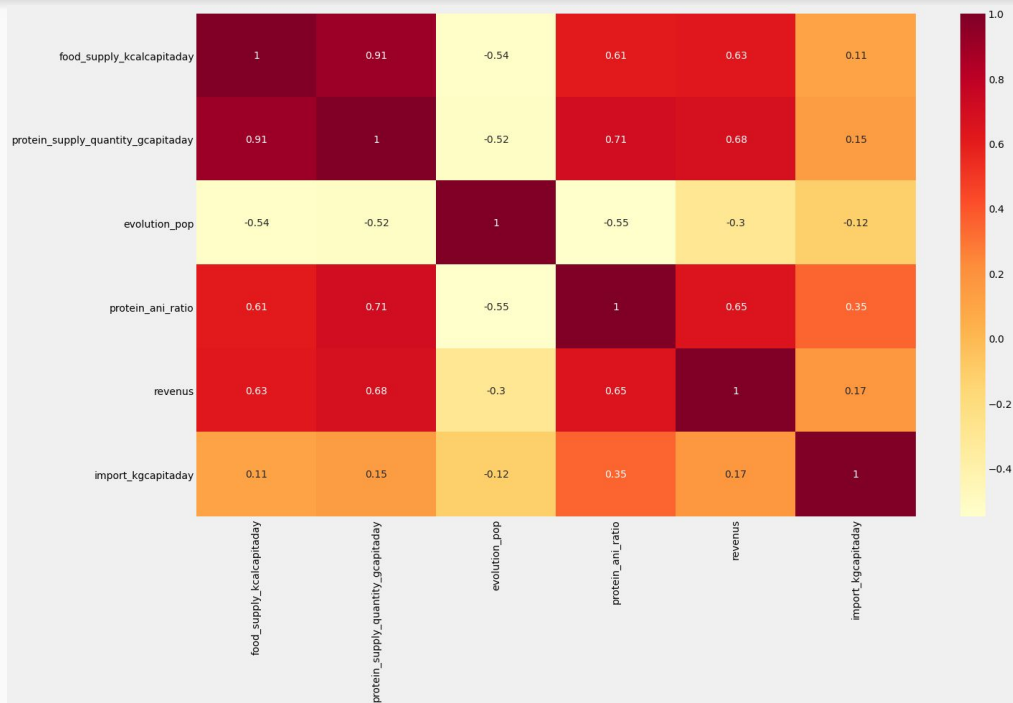
# MATRICE PAR PAIRES



- Corrélations
- Distributions asymétriques
- Comparaisons



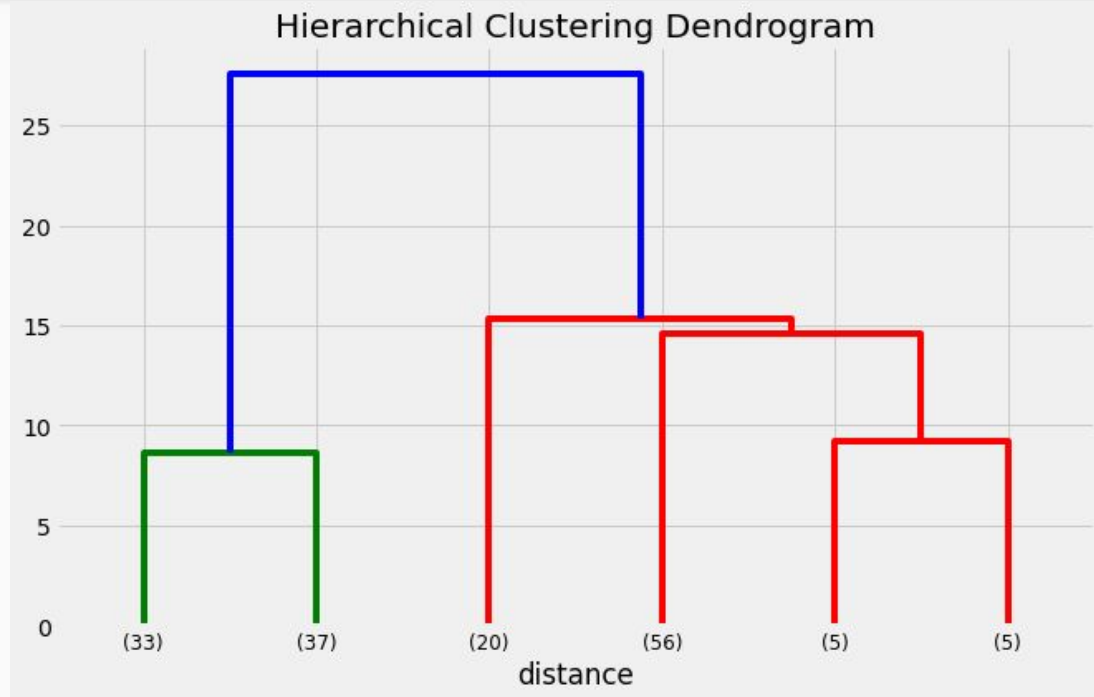
# MATRICE CORRELATIONS



- Corrélations
- Classification

# **CLASSIFICATION HIÉRARCHIQUE**

# DENDROGRAMME

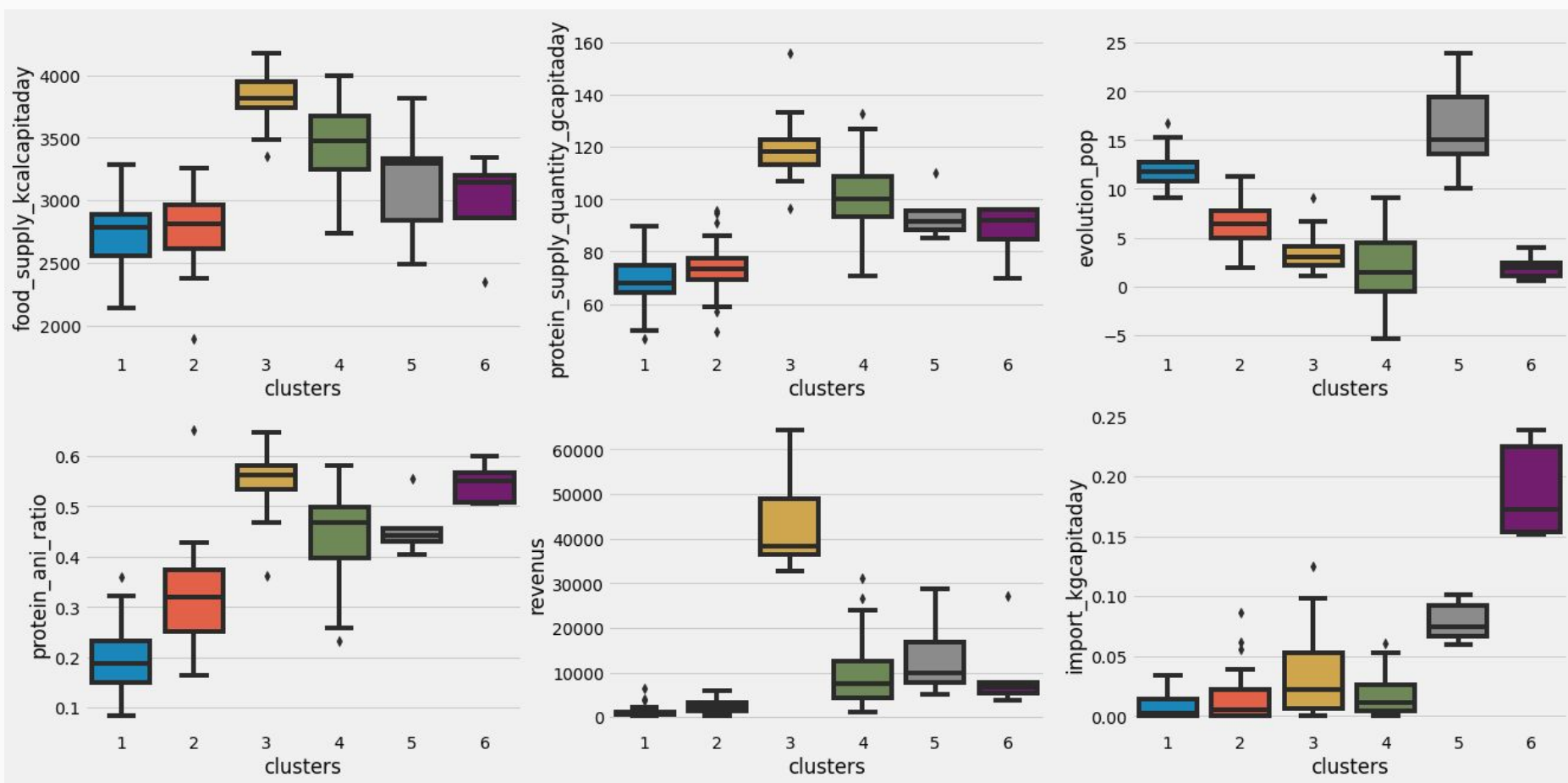


# DESCRIPTION CLUSTERS

	food_supply_kcalcapitaday	protein_supply_quantity_gcapitaday	evolution_pop	protein_ani_ratio	revenus	import_kgcapitaday
clusters						
1	2718.625279	68.665095	12.093939	0.194681	1128.737339	0.008124
2	2786.797487	73.619680	6.275676	0.314920	2477.764935	0.014275
3	3831.010826	118.974844	3.390000	0.550305	43635.705925	0.033655
4	3465.835959	99.950891	1.832143	0.442884	9331.712691	0.016289
5	3155.832667	94.104296	16.400000	0.457690	13611.744393	0.078670
6	2980.432667	87.760296	2.000000	0.546401	10125.944530	0.187987

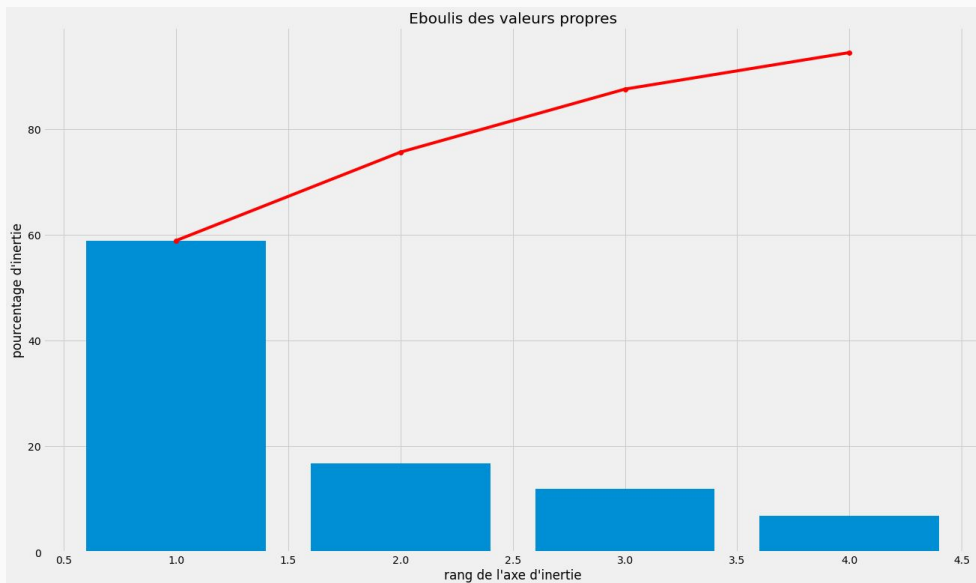
**Cluster 3**

France (20 pays)



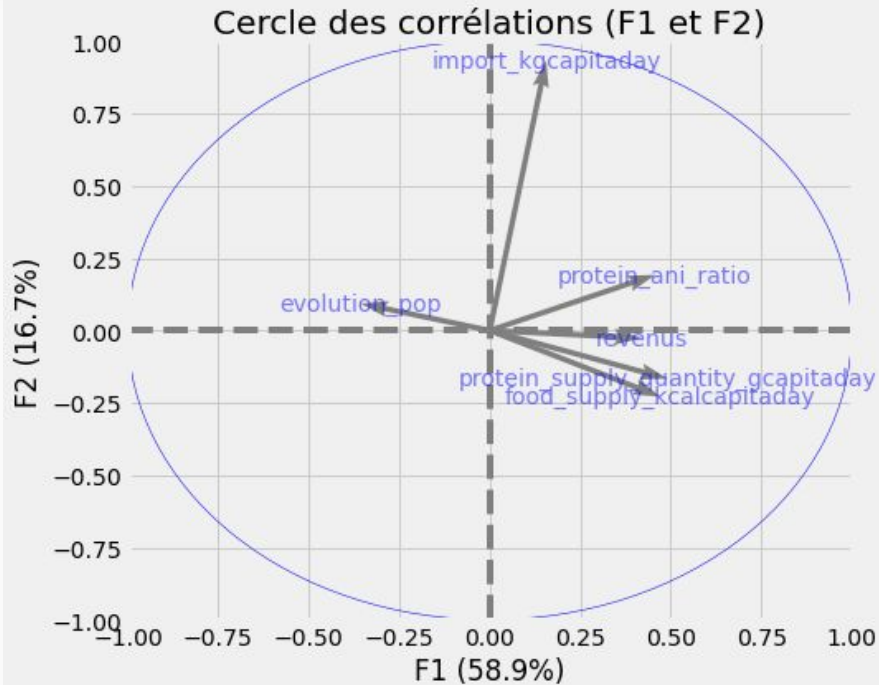
**ACP**

# EBOULIS VALEURS PROPRES



Choix de conserver composante 1  
(80%, méthode coude - critère de  
Kaiser,  $100/p$ , soit 50%)

# CERCLE CORRÉLATIONS 1er PLAN



F1 : relation population / alimentation

**Disponibilité**

0,47

**Protéines**

0,49

**Ratio**

0,45

**Population**

-0,35

**Revenus**

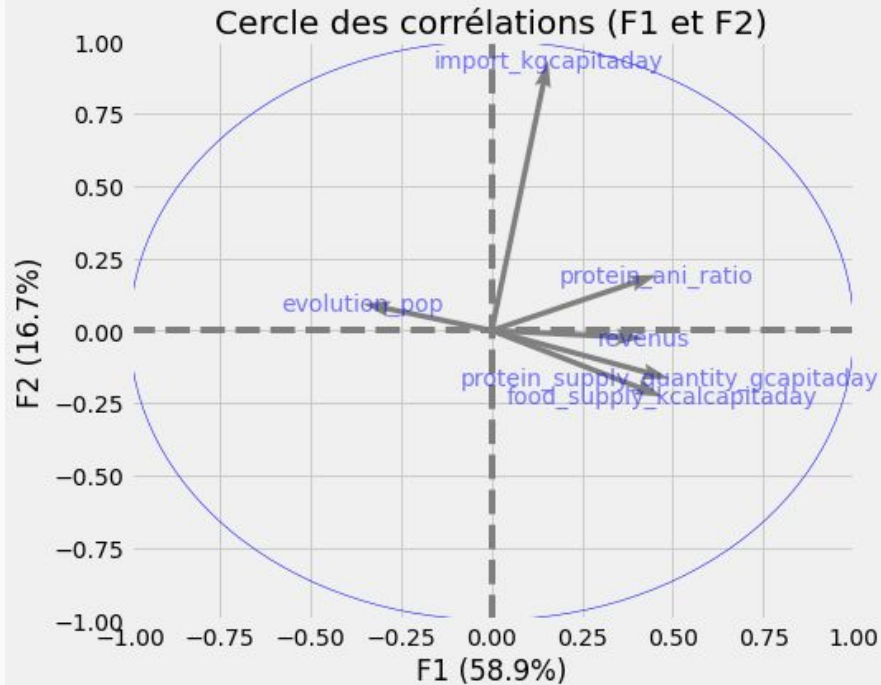
0,41

**Import**

0,15



# CERCLE CORRÉLATIONS 1er PLAN



F2 : Ensemble variables

**Disponibilité**

-0,22

**Protéines**

-0,16

**Ratio**

0,19

**Population**

0,09

**Revenus**

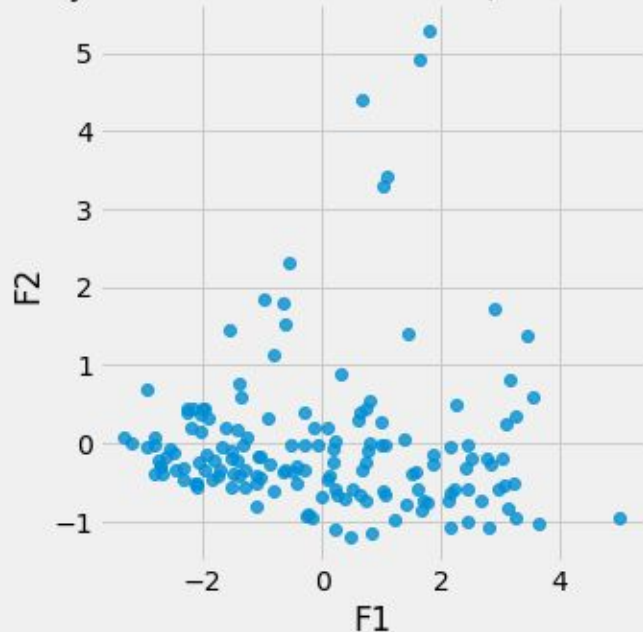
-0,02

**Import**

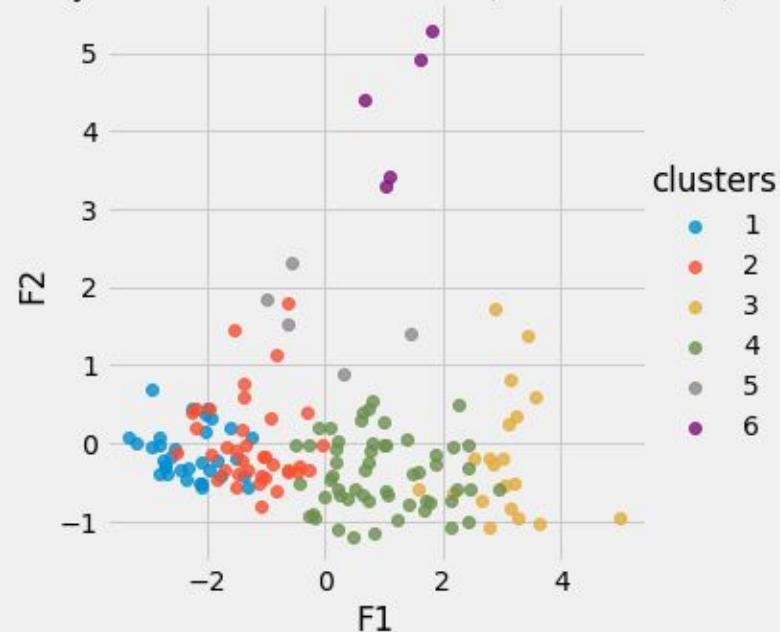
0,93

# PROJECTION DES INDIVIDUS ET CLUSTERS

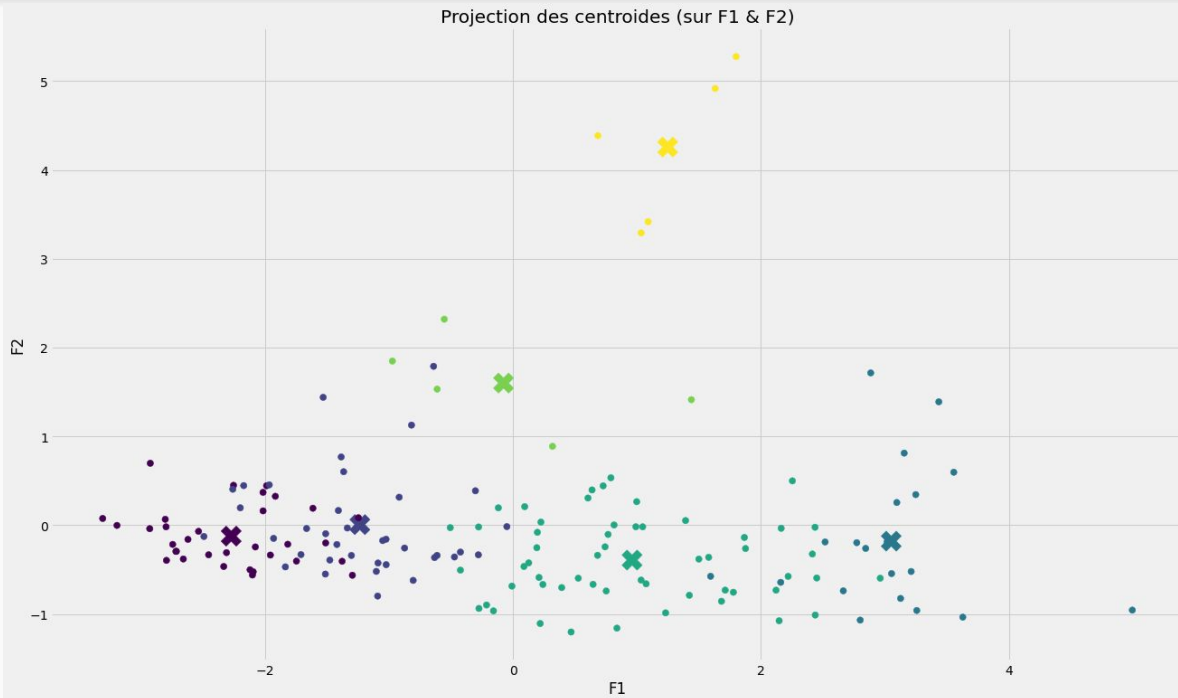
Projection des individus (sur F1 & F2)



Projection des individus (sur F1 & F2)



# PROJECTION CENTROÏDES



clusters	F1	F2
1	-2.272011	-0.122557
2	-1.234454	0.007248
3	3.050207	-0.178747
4	0.960500	-0.391912
5	-0.075960	1.600955
6	1.247765	4.258682

# COMPARAISONS VARIABLES

## **Adhésion loi normale :**

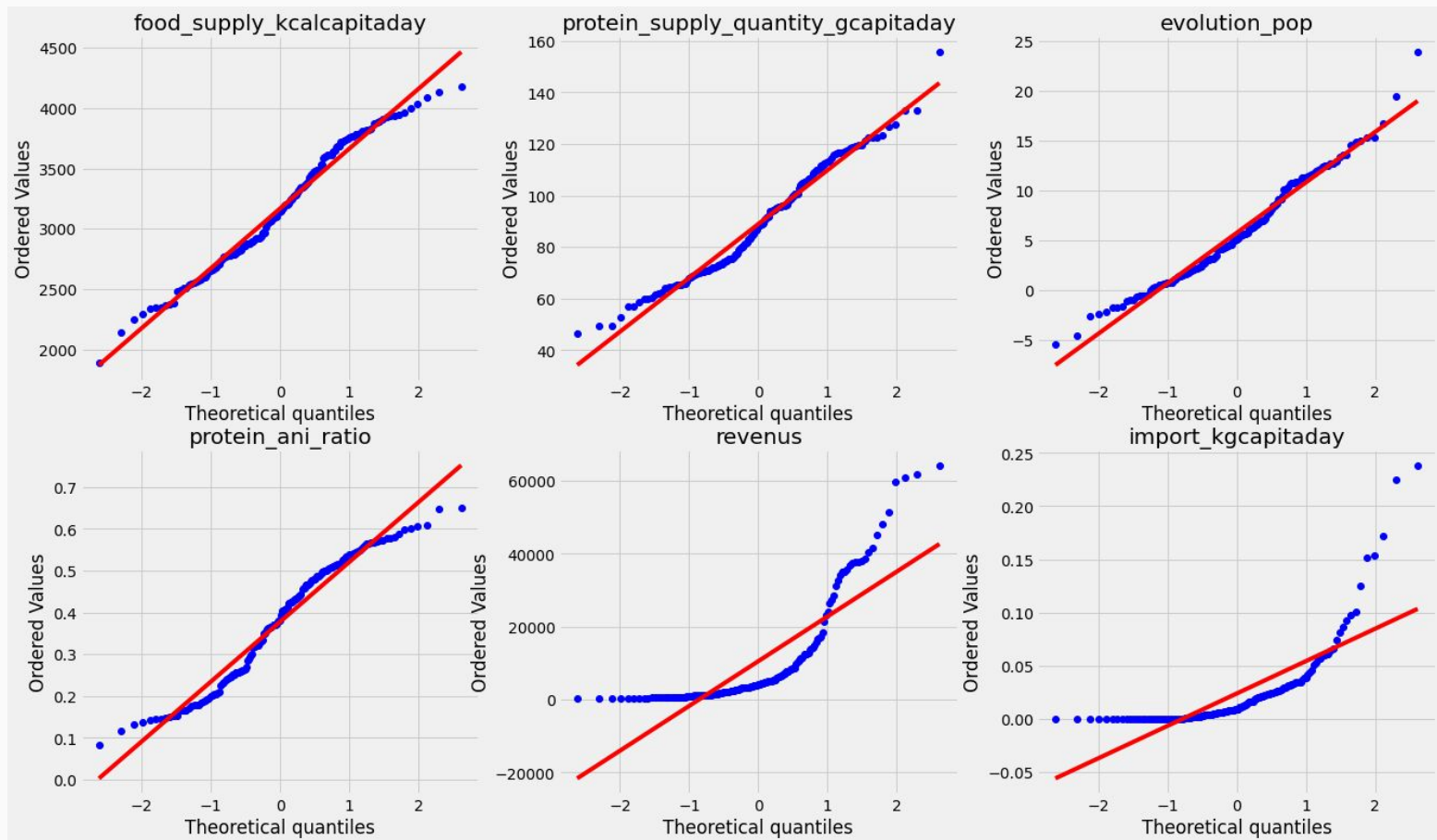
- (Histogramme)
- Shapiro-Wilk
- Droite de Henry

## **Egalité de variance :**

- Bartlett

## **Comparaison moyenne :**

- t-test



# Adhésion loi normale

## Shapiro-wilk

H0 : adhésion loi normale  
H1 : pas adhésion loi normale  
Seuil : 2%

**Disponibilité**     $p = 0,021$

**Population**     $p = 0,008$

**Protéines**     $p = 0,005$

**Revenus**     $p = 1,6e-16$

**Ratio**     $p = 7,07e-5$

**Import**     $p = 1,9e-18$

On ne peut pas rejeter l'hypothèse de normalité au niveau de test 2% pour l'ensemble des variables, sauf pour la disponibilité alimentaire.

## Tests comparaisons (cluster 3 & 4)

### Variance

H0 : égalité variance

H1 : pas d'égalité variance

Seuil : 5%

**Disponibilité**

$p = 0,08$

### Moyenne

H0 : égalité moyenne

H1 : pas d'égalité moyenne

Seuil : 5%

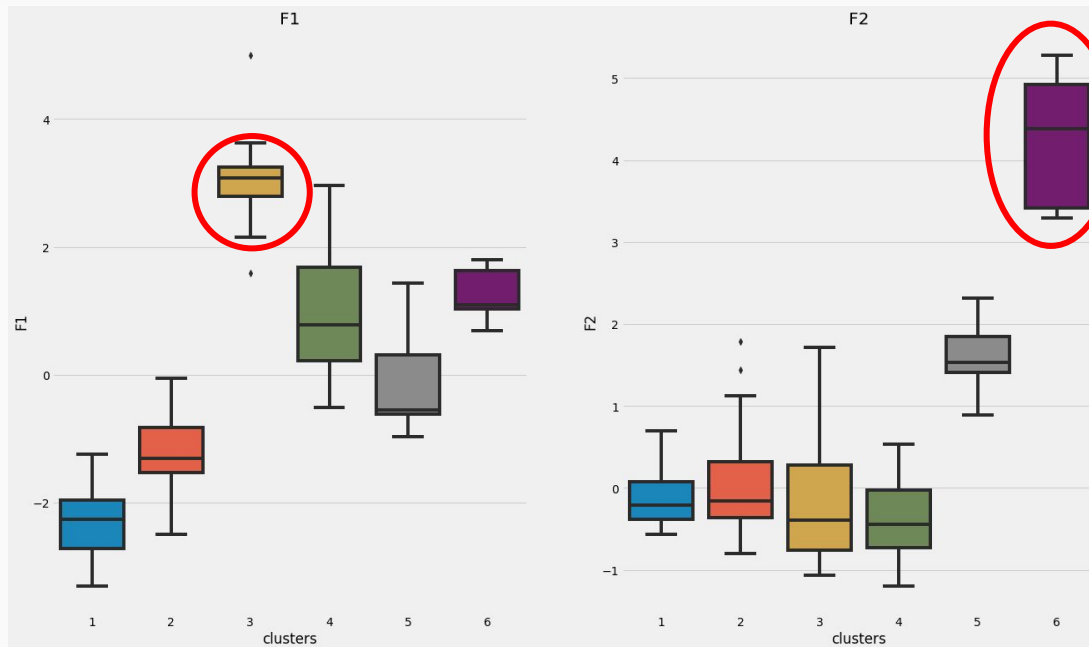
**Disponibilité**

$p = 1,5e-6^*$

**On ne peut pas rejeter l'hypothèse d'égalité de variance au niveau de test 5% pour l'ensemble des variables.**

**On peut rejeter l'hypothèse d'égalité de moyenne au niveau de test 5%.**

# CARACTÉRISATION CLUSTERS



- Cluster 3 le plus intéressant.
- Cluster 6 pour import



# OPPORTUNITÉS

# PROBLÉMATIQUE

## **Problématique :**

1. production en France pour exportation produits frais.
2. instabilité politique et commerciale GB lié au Brexit

# RECOMMENDATIONS

**Recommandations :** Éviter cluster 6

## Cluster 6

- Bahamas
- Dominique
- Saint-Vincent-et-les Grenadines
- Sainte-Lucie
- Samoa

# RECOMMANDATIONS

## Recommandations : cluster 3

Sans pays hors Europe

- Australie
- Canada
- Israël
- Nouvelle-Zélande
- Émirats arabes unis
- États-Unis d'Amérique

et sans pays U.E.

- “Irlande”
- Islande
- Norvège
- “Royaume-Uni”

# RECOMMENDATIONS

## Recommandations : cluster 3

1

- Allemagne
- Luxembourg
- Belgique
- Suisse

2

- Pays-Bas
- Autriche

3

- Danemark
- Finlande
- Suède

# MERCI !



@xavbarbier



<https://www.linkedin.com/in/barbierxavier/>



<https://github.com/xavierbarbier/>



[contact@xavierbarbier.com](mailto:contact@xavierbarbier.com)