



La poule qui chante

Développement a l'international : Etude de marché

Jorge Ocon Garcia



Sommaire

- Le contexte
- Objectif
- L'analyse en composants principaux
- Clustering Méthode 1
- Clustering Méthode 2
- Conclusion



Le contexte



Dans un objectif de développement, La poule qui chante, une entreprise française d'agroalimentaire, souhaite étendre son marché à international.

L'objectif

Proposer des groupements de pays que l'on pourra cibler pour exporter nos poulets



La poule qui chante



L'analyse en composants principaux

1. Scaler les variables

L'objectif principal de cette étape est de établir les liaisons entre variables et la variabilité entre les individus

Scale variables

```
Entrée [20]: scaler = StandardScaler()
```

```
Entrée [21]: #scaler.fit(X)
```

```
Entrée [22]: #population_PIB_dispo_alimentaire_scaled = scaler.transform(X)
```

```
Entrée [23]: X_scaled = scaler.fit_transform(X)
```

```
Entrée [24]: population_PIB_dispo_alimentaire_scaled = pd.DataFrame(X_scaled)
population_PIB_dispo_alimentaire_scaled.columns = population_PIB_dispo_alimentaire.columns
```

```
Entrée [25]: idx = ["mean", "std"]

pd.DataFrame(population_PIB_dispo_alimentaire_scaled).describe().round(2).loc[idx, :]
```

Out[25]:

	Disponibilité alimentaire en quantité (kg/personne/an)	Disponibilité intérieure	Exportations - Quantité	Importations - Quantité	Production	2017 (en milliers)	Valeur US \$	Valeur US \$ par habitant
mean	-0.0	-0.0	0.0	-0.0	-0.0	-0.0	0.0	0.0
std	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

L'analyse en composants principaux

ACP

```
Entrée [26]: n_components = 6  
pca = PCA(n_components=n_components)
```

```
Entrée [27]: pca.fit(population_PIB_dispo_alimentaire_scaled)
```

```
Out[27]: PCA(n_components=6)
```

```
Entrée [28]: pca.explained_variance_ratio_
```

```
Out[28]: array([0.5021144 , 0.18749145, 0.11437595, 0.07940166, 0.0762646 ,  
               0.03142656])
```

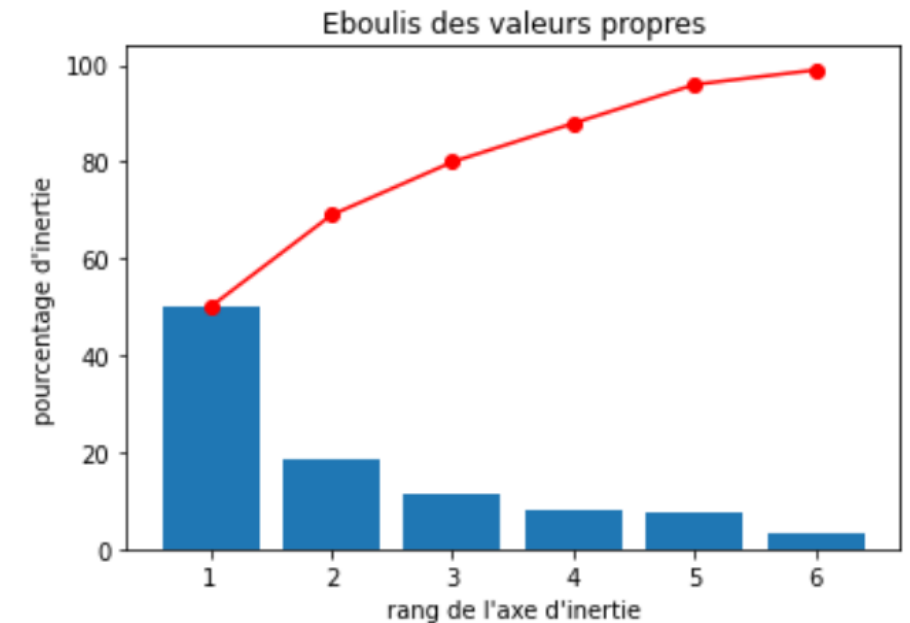
```
Entrée [29]: scree = (pca.explained_variance_ratio_*100).round(2)  
scree
```

```
Out[29]: array([50.21, 18.75, 11.44,  7.94,  7.63,  3.14])
```

```
Entrée [30]: scree_cum = scree.cumsum().round()  
scree_cum
```

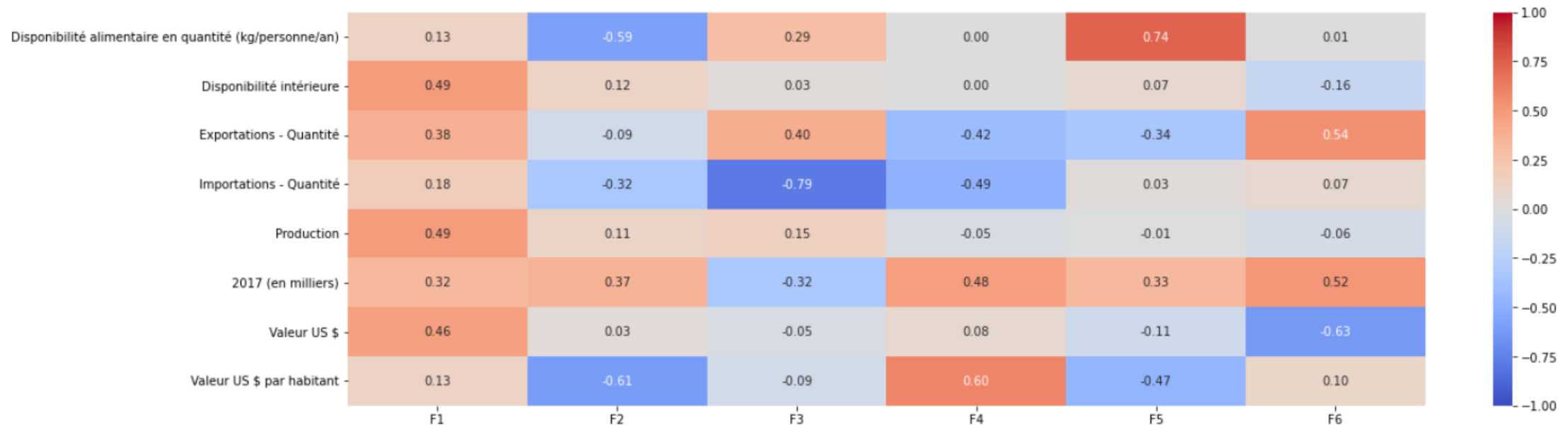
```
Out[30]: array([50., 69., 80., 88., 96., 99.])
```

2. Choisir le nombre de composants



L'analyse en composants principaux

3. Contribution de chaque variable a chaque composant



Clustering Méthode 1

Classification ascendant hiérarchique

La classification hiérarchique permet de créer un **arbre** qui regroupe les différents individus de façon plus ou moins forte en fonction de la **profondeur** choisie.

1. Choisir le méthode de linkage

Cluster ascendant hierarchique

```
Entrée [41]: Z = linkage(X_scaled, method="ward")  
pd.DataFrame(Z)
```


Clustering Méthode 1



La poule qui chante

3. Le silhouette score

```
Entrée [44]: plt.style.use("fivethirtyeight")  
plt.plot(range(2, 16), silhouette_coefficients)  
plt.xticks(range(2, 16))  
plt.xlabel("Number of Clusters")  
plt.ylabel("Silhouette Coefficient")  
plt.show()
```

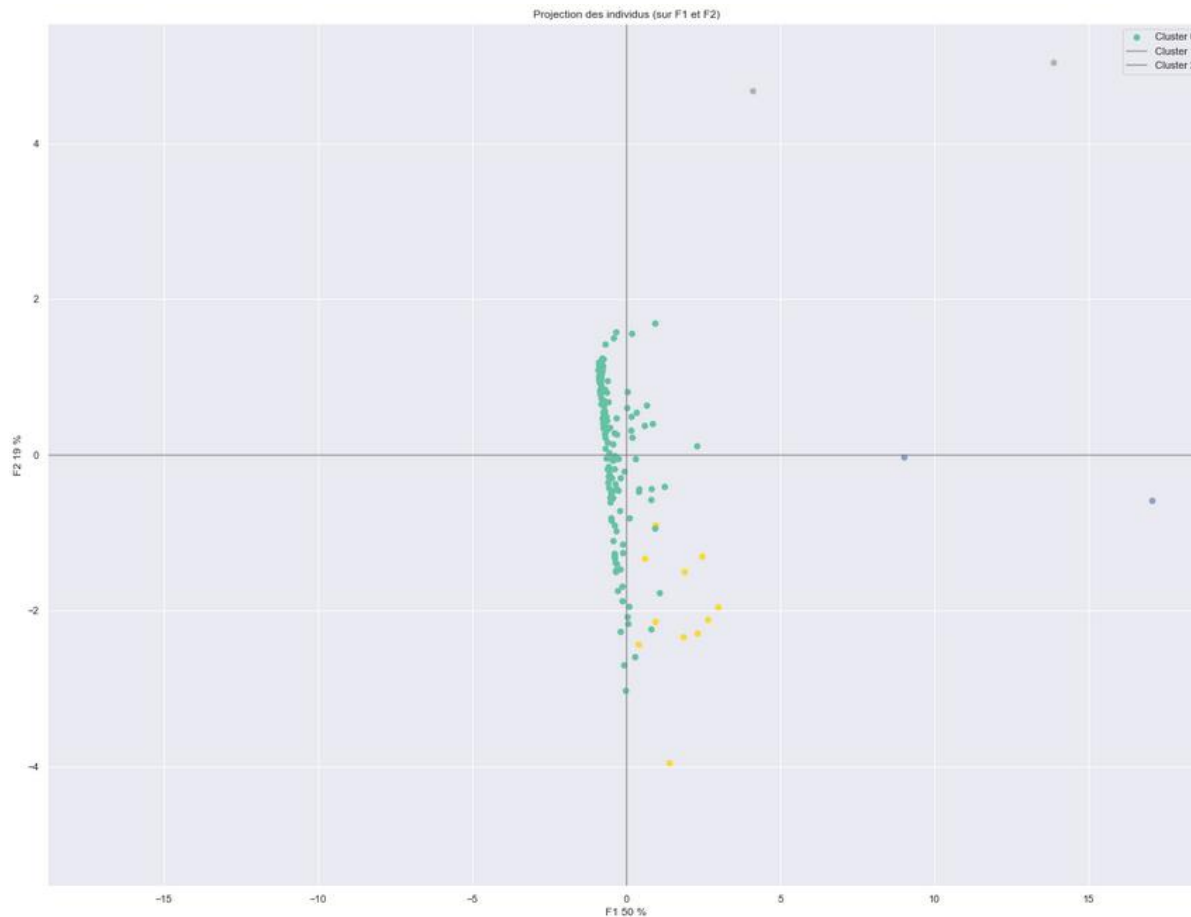


Clustering Méthode 1



La poule qui chante

4. Représentation des clusters sur les composants principaux



Clustering Méthode 2

K-means

L'algorithme du k-means travaille avec les centres de gravité des groupes, son objectif est de trouver des groupes en faisant en sorte de minimiser l'inertie intraclasse

1. Choisir le nombre de cluster

```
Entrée [97]: kmeans = KMeans(n_clusters=4)
```

```
Entrée [98]: kmeans.fit(X_scaled)
```

```
Out[98]: KMeans(n_clusters=4)
```

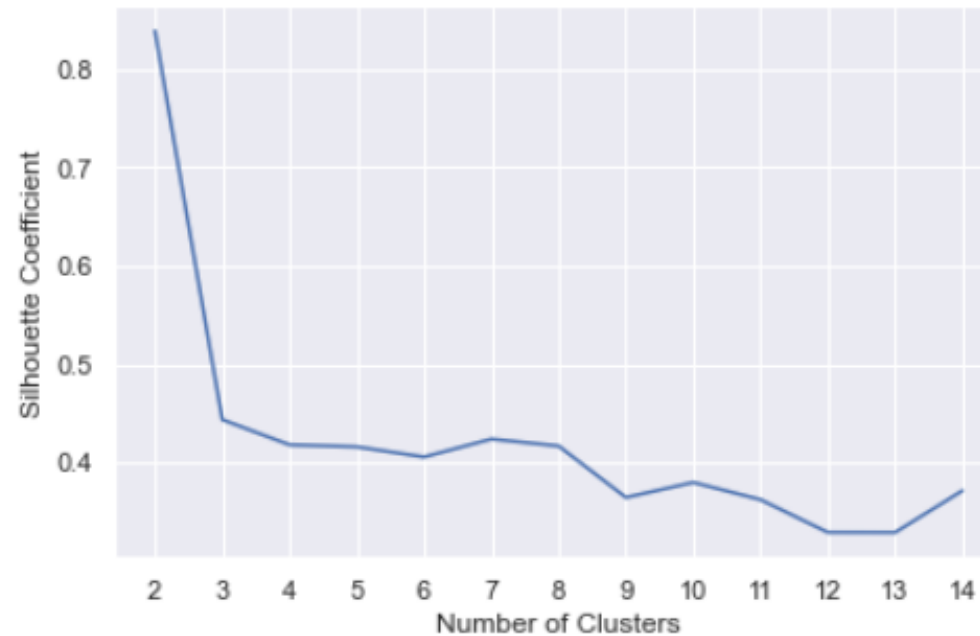
Clustering Méthode 2



La poule qui chante

2. Le silhouette score

```
Entrée [56]: plt.plot(range(2, 15), km_silhouette_coefficients)
plt.xticks(range(2, 15))
plt.xlabel("Number of Clusters")
plt.ylabel("Silhouette Coefficient")
plt.show()
```

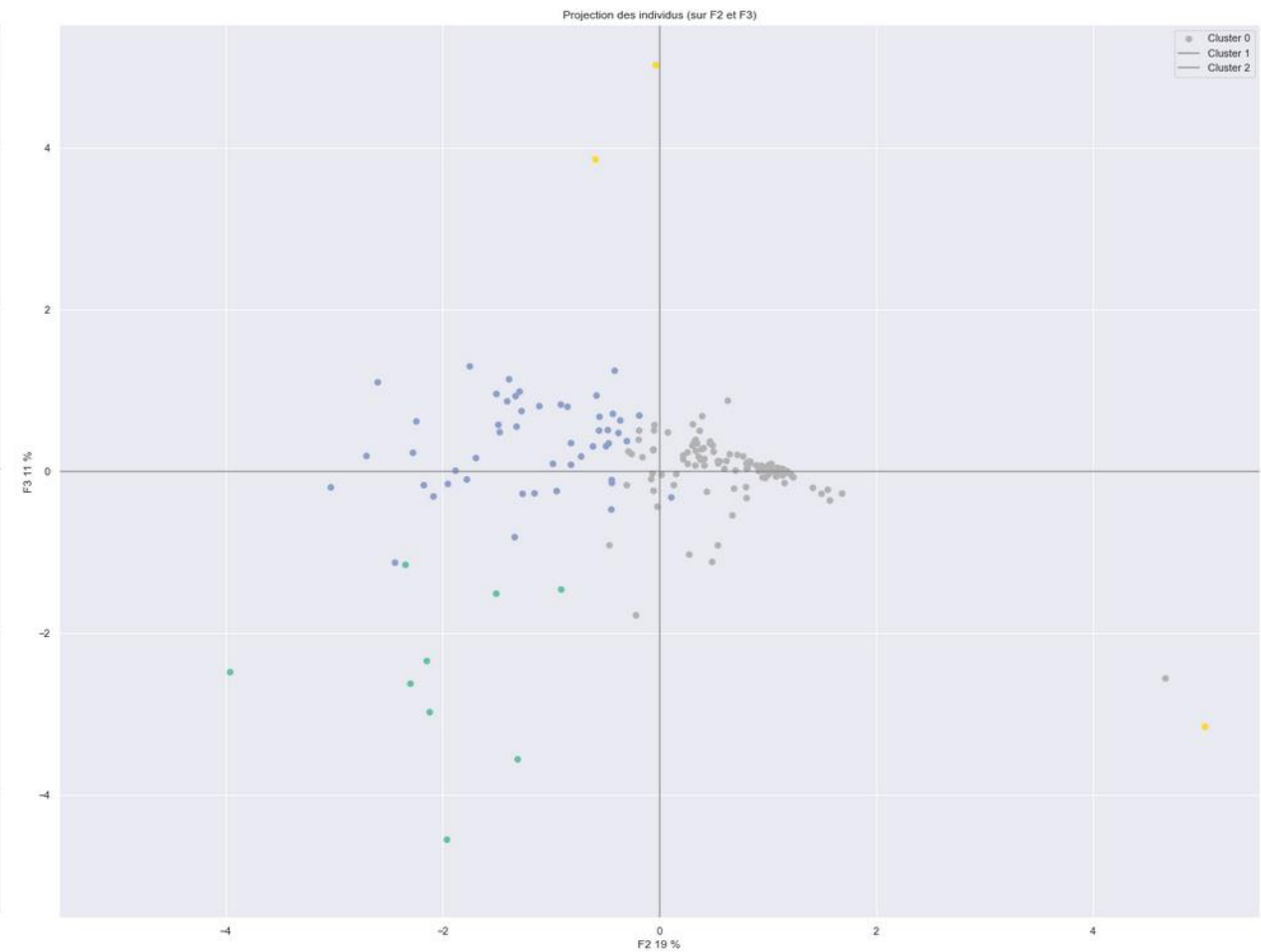
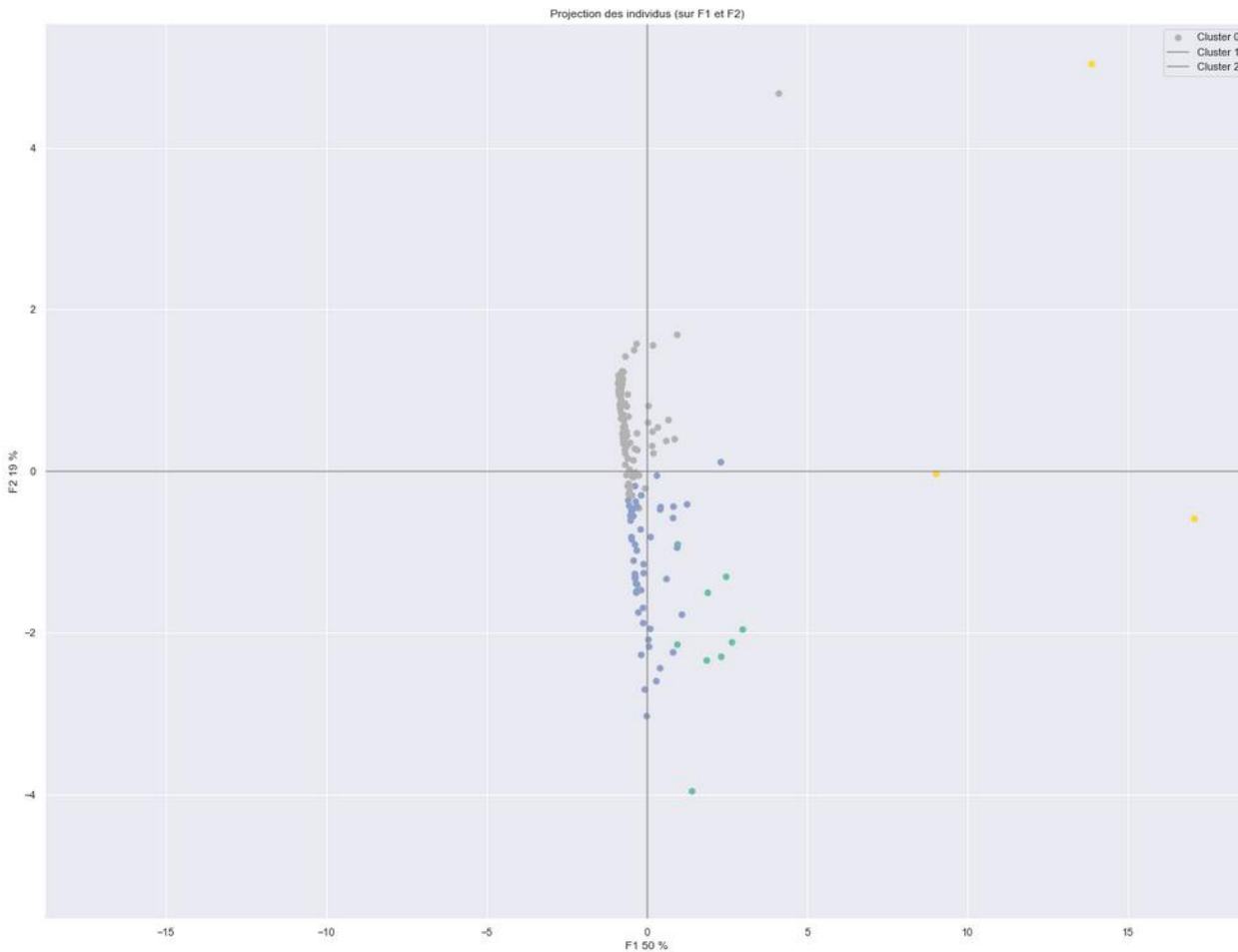


Clustering Méthode 2



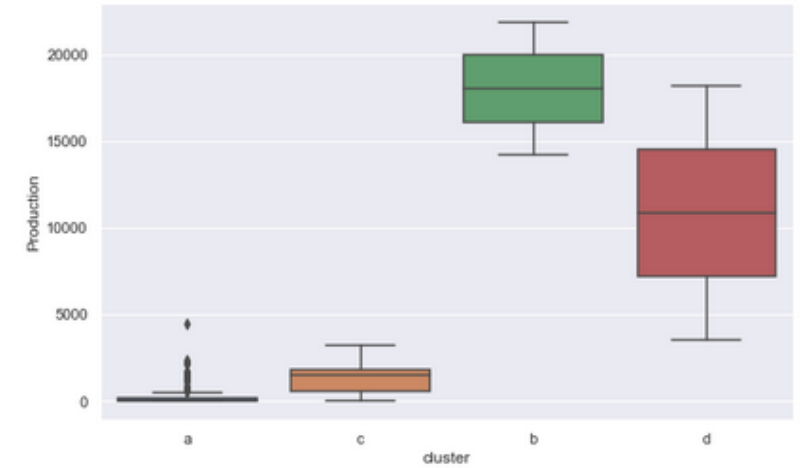
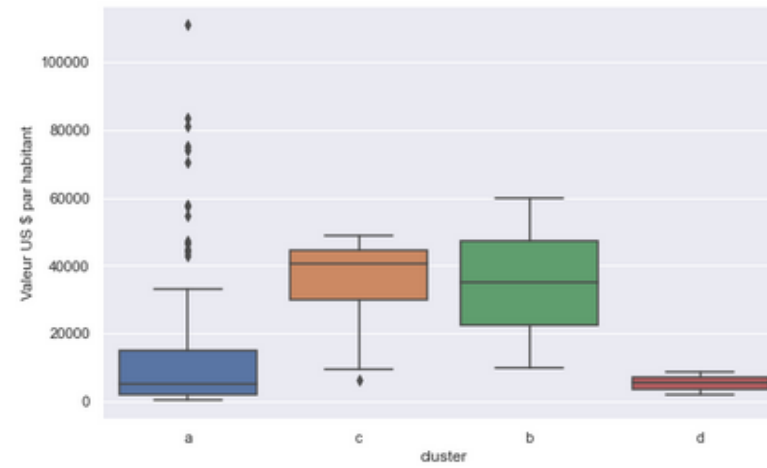
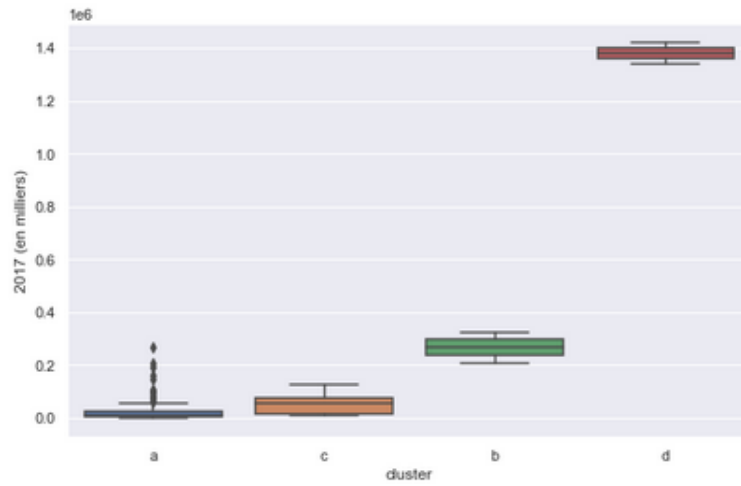
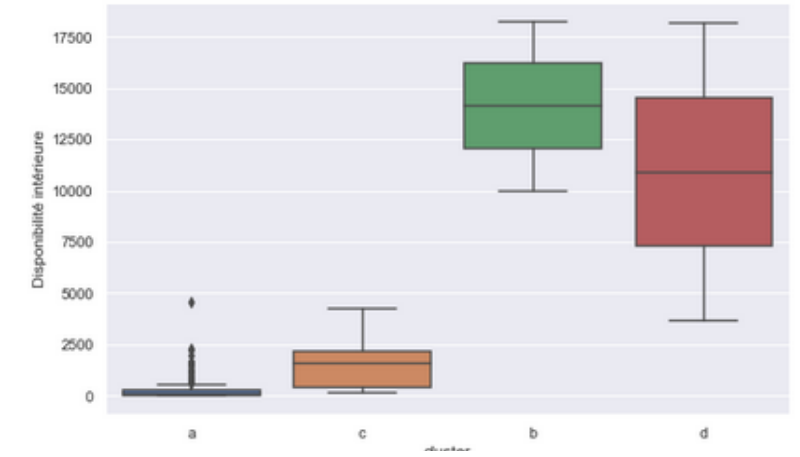
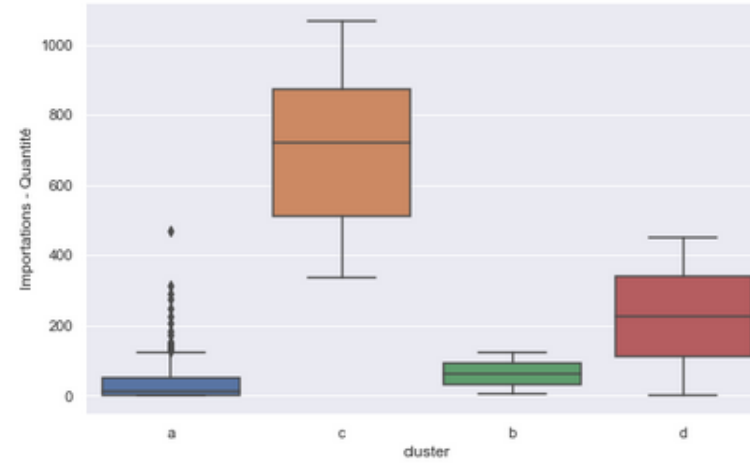
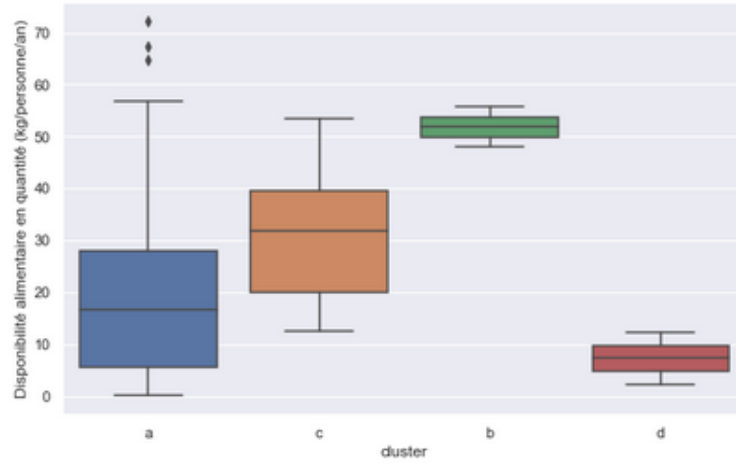
La poule qui chante

3. Représentation des clusters sur les composants principaux



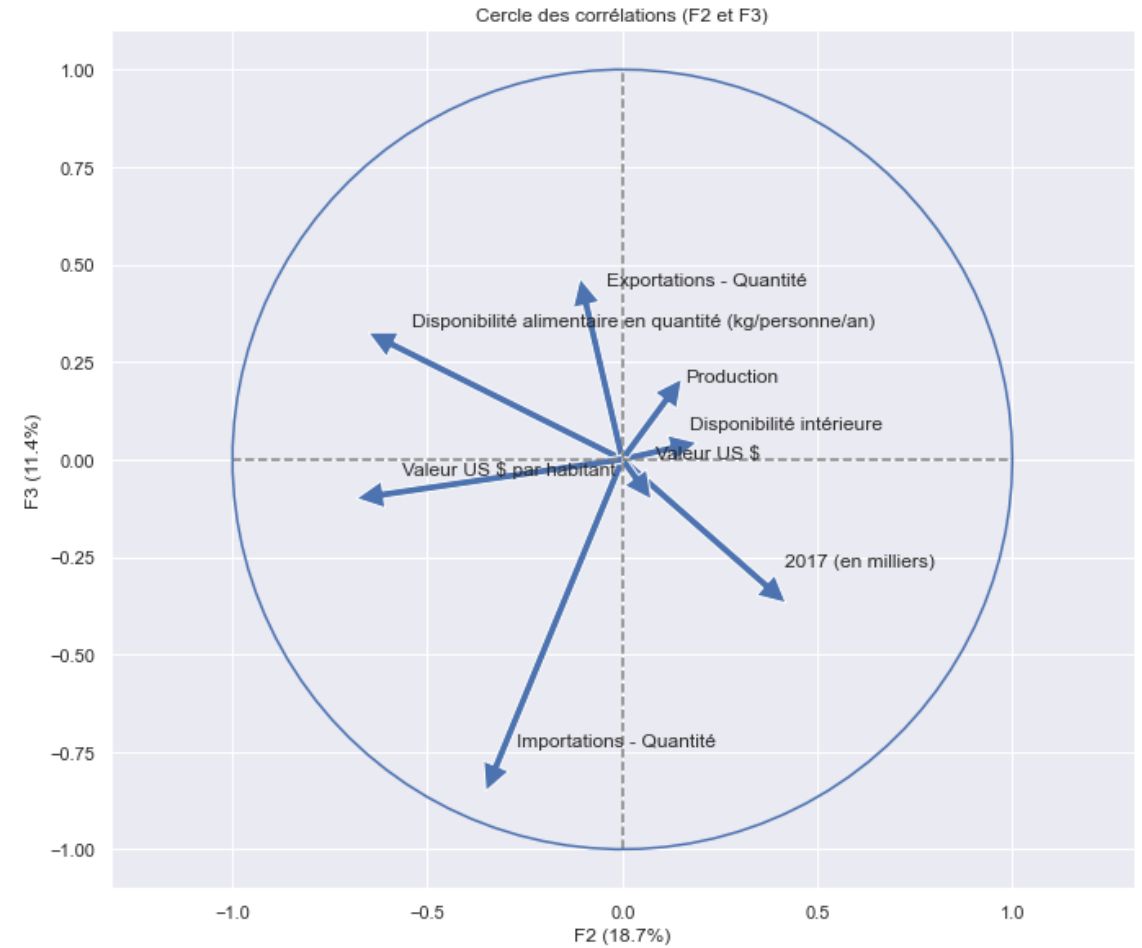


Resultats





Resultats





Resultats

Entrée [54]: `population_PIB_dispo_alimenature[population_PIB_dispo_alimenature['cluster'] == 'c']`

Out[54]:

Zone	Disponibilité alimentaire en quantité (kg/personne/an)	Disponibilité intérieure	Exportations - Quantité	Importations - Quantité	Production	2017 (en milliers)	Valeur US \$	Valeur US \$ par habitant	cluster
Afrique du Sud	35.69	2118.0	63.0	514.0	1667.0	57009.756	3.490067e+05	6121.876572	c
Allemagne	19.47	1739.0	646.0	842.0	1514.0	82658.409	3.690849e+06	44651.829102	c
Arabie saoudite	43.36	1435.0	10.0	722.0	616.0	33101.179	6.885861e+05	20802.464299	c
Belgique	12.65	152.0	656.0	338.0	463.0	11419.748	5.027647e+05	44025.903247	c
Chine - RAS de Hong-Kong	53.51	280.0	663.0	907.0	24.0	7306.322	3.412420e+05	46705.027813	c
France	22.90	1573.0	501.0	506.0	1750.0	64842.509	2.595151e+06	38720.645922	c
Japon	18.50	2415.0	10.0	1069.0	2215.0	127502.725	4.930837e+06	38672.407740	c
Mexique	32.52	4219.0	9.0	972.0	3249.0	124777.324	1.158912e+06	9287.841557	c
Pays-Bas	20.33	372.0	1418.0	608.0	1100.0	17021.347	8.338696e+05	48989.638816	c
Royaume-Uni de Grande-Bretagne et d'Irlande du Nord	31.94	2234.0	359.0	779.0	1814.0	66727.461	2.699017e+06	40448.365256	c
Émirats arabes unis	43.47	412.0	94.0	433.0	48.0	9487.203	3.856055e+05	40644.804043	c