DPENCLASSROOMS





Etude du marché du poulet au niveau international

Projet 5: Produisez une étude de marché











Contexte

☐ Etude du marché de **poulet** d'une entreprise agroalimentaire



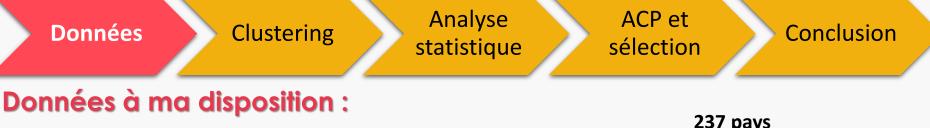
Entreprise agroalimentaire qui souhaite se développer à l'international

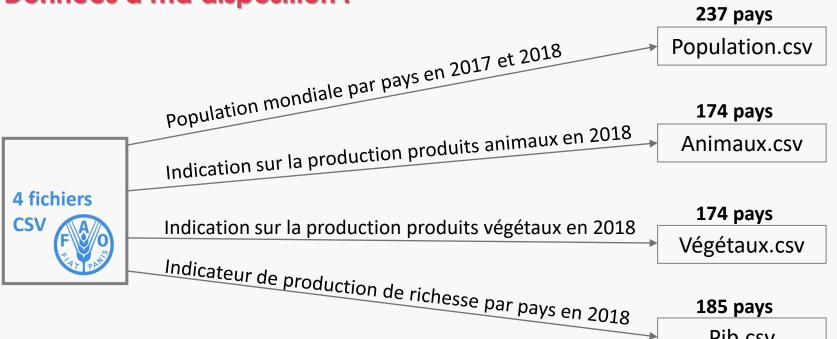


- Stratégie : exporter les produits dans les pays propices à une insertion de marché du poulet
- Analyse des caractéristiques pays (différence population, régimes alimentaires, PIB..)

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

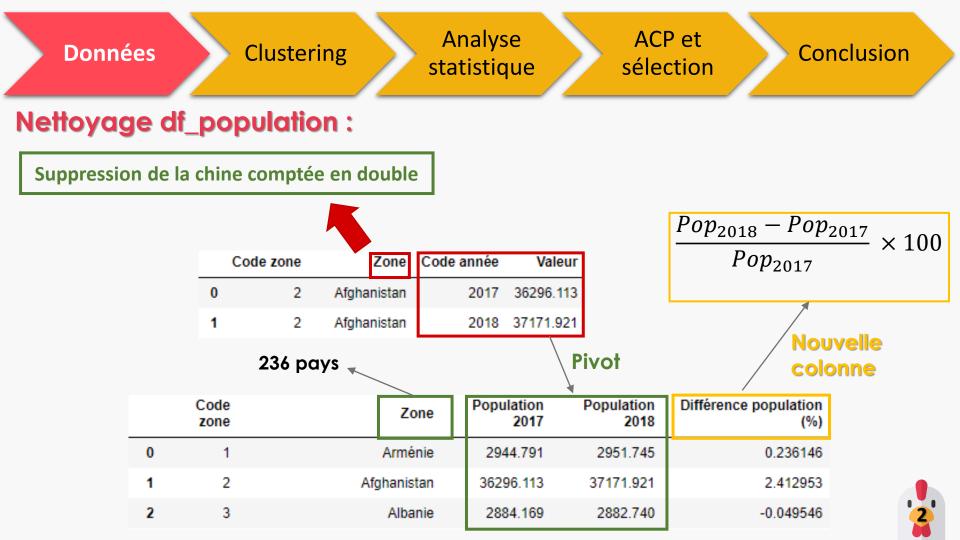


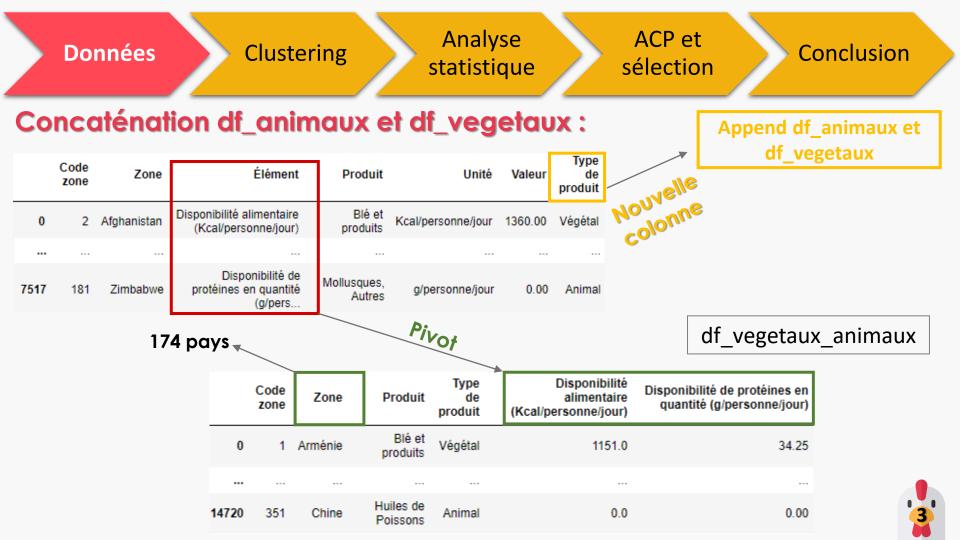












Clustering

Analyse statistique

ACP et sélection

Conclusion

Merge df_population et df_vegetaux_animaux :

		Code zone			Zone	Population 2017	Рорі	ulation 2018	Différence population (%)
0		1		A	rménie	2944.791	29	51.745	0.236146
1		2		Afgh	anistan	36296.113	371	71.921	2.412953
2		3			Albanie	2884.169	28	82.740	-0.049546
		Code		Produit	Type de produit	•	nibilité entaire ne/jour)		ibilité de protéines en ntité (g/personne/jour)
	0	1	Arménie	Blé et produits	Végétal		1151.0		34.25
	1	1	Arménie	Orge et produits	Végétal		6.0		0.20

df_population

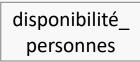


df_vegetaux animaux

173 pays

	Code zone	Zone	Population 2017	Population 2018	Différence population (%)	Produit	Type de produit	Disponibilité alimentaire (Kcal/personne/jour)	Disponibilité de protéines en quantité (g/personne/jour)
0	1	Arménie	2944.791	2951.745	0.236146	Blé et produits	Végétal	1151.0	34.25
1	1	Arménie	2944.791	2951.745	0.236146	Orge et produits	Végétal	6.0	0.20

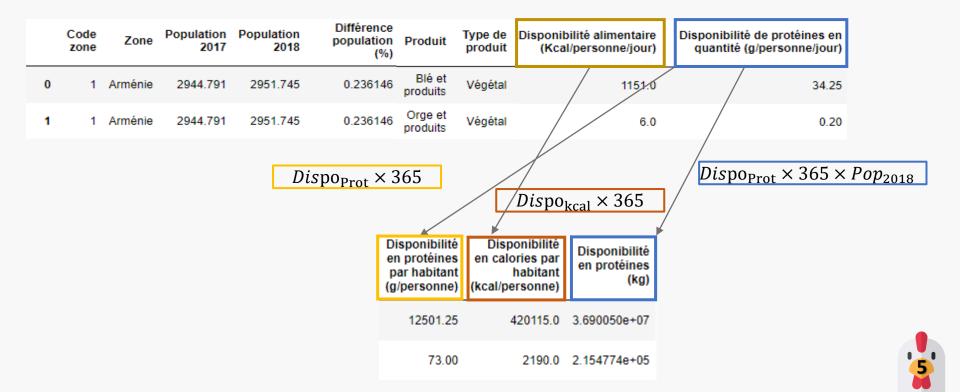






Données Clustering Analyse ACP et sélection Conclusion

Création de différente colonne :





	zone	Zone	2018	population (%)	Produit	produit	Annee	par habitant (g/personne)	habitant (kcal/personne)	protéines (kg)
56	1	Arménie	2951.745	0.236146	Viande de Bovins	Animal	2018	2752.10	40150.0	8.123497e+06
57	1	Arménie	2951.745	0.236146	Viande d'Ovins/Caprins	Animal	2018	390.55	7665.0	1.152804e+06
58	1	Arménie	2951.745	0.236146	Viande de porcins	Animal	2018	737.30	19710.0	2.176322e+06
59	1	Arménie	2951.745	0.236146	Viande de Volailles	Animal	2018	1872.45	18615.0	5.526995e+06
						_				



disponibilité personnes

Conclusion

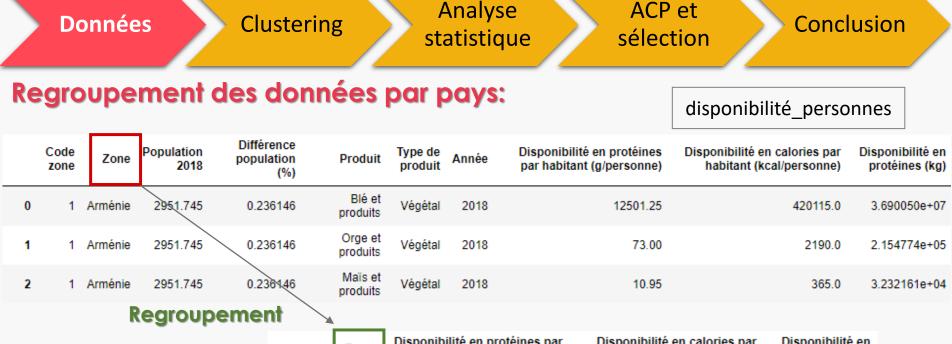
Disponibilité en

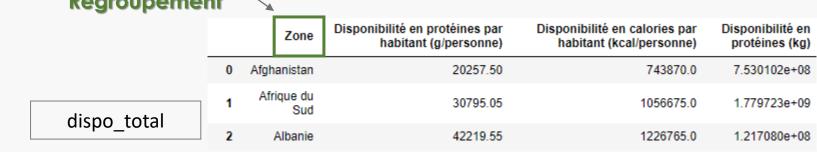


ACP et

df volailles

Disponibilité en protéines de volailles (kg) Arménie 59 5.526995e+06 Afghanistan 6.919553e+06 210 Albanie 3.703744e+06







Analyse statistique

ACP et sélection

Conclusion

Regroupement des données par produits végétal et animal :



disponibilité personnes

Regroupement

	Code zone	Zon	Type e de produit	2018	Différence population (%)	Disponibilité en protéines (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)
0	1	Armén	ie Animal	2951.745	0.236146	4.881640e+07	16538.15	290175.0
1	1	Armén	ie Végétal	2951.745	0.236146	5.278119e+07	17881.35	801905.0

dispo pays

Pivot

	Code zone	Zone	Population Différenc 2018 population (%		Dispo protéines animale (kg)	Dispo protéines végétale (kg)
0	1	Arménie	2951.745	0.236146	4.881640e+07	5.278119e+07
1	2	Afghanistan	37171.921	2.412953	1.463960e+08	6.066142e+08



Analyse statistique

ACP et sélection

Conclusion

Merge des dataframes dispo_pays et dispo_total :

			_			
	Code zone	Zone	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)
0	1	Arménie	2951.745	0.236146	4.881640e+07	5.278119e+07
1	2 Afg	ghanistan	37171.921	2.412953	1.463960e+08	6.066142e+08
	Zone	Dispon	ibilité en protéi habitant (g/pe		nibilité en calories par bitant (kcal/personne)	Disponibilité en protéines (kg)
0	Zone Afghanistan		habitant (g/pe			•

dispo_pays



dispo_total



Jointure de type inner sur zone

dispo_pays_habitants

	Code zone	Zone	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Disponibilité en protéines (kg)
0	1	Arménie	2951.745	0.236146	4.881640e+07	5.278119e+07	34419.50	1092080.0	1.015976e+08
1	2	Afghanistan	37171.921	2.412953	1.463960e+08	6.066142e+08	20257.50	743870.0	7.530102e+08
2	3	Albanie	2882.740	-0.049546	6.496283e+07	5.674515e+07	42219.55	1226765.0	1.217080e+08



Analyse statistique

ACP et sélection

Conclusion

Merge des dataframes dispo_pays_habitant et df_volailles :

	Code zone	Zone	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Disponibilité en protéines (kg)
0	1	Arménie	2951.745	0.236146	4.881640e+07	5.278119e+07	34419.50	1092080.0	1.015976e+08
1	2	Afghanistan	37171.921	2.412953	1.463960e+08	6.066142e+08	20257.50	743870.0	7.530102e+08
2	3	Albanie	2882.740	-0.049546	6.496283e+07	5.674515e+07	42219.55	1226765.0	1.217080e+08

 Zone
 Disponibilité en protéines de volailles (kg)

 59
 Arménie
 5.526995e+06

 143
 Afghanistan
 6.919553e+06

 210
 Albanie
 3.703744e+06

dispo_pays_ habitant



df_volailles

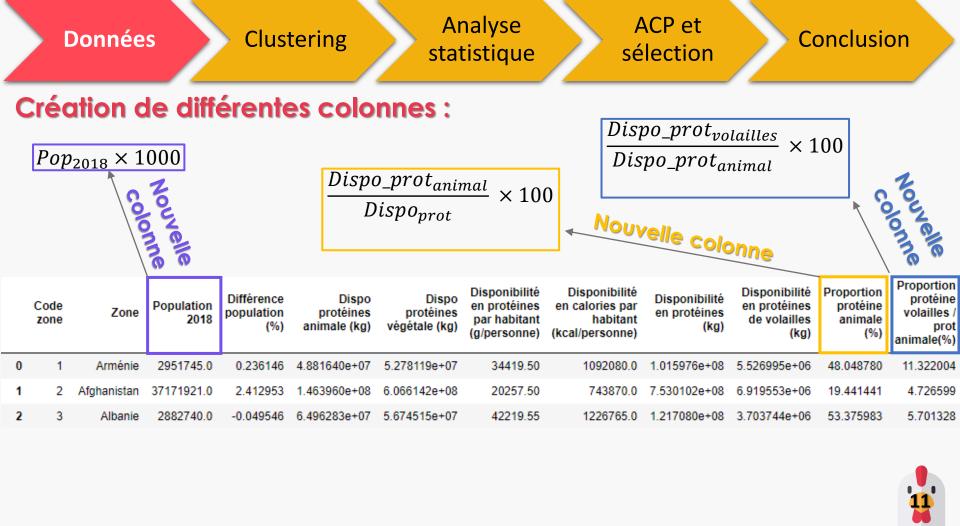
Jointure de type inner sur zone



dispo_pays_volailles

Coo		Zone	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Disponibilité en protéines (kg)	Disponibilité en protéines de volailles (kg)	
0	1	Arménie	2951.745	0.236146	4.881640e+07	5.278119e+07	34419.50	1092080.0	1.015976e+08	5.526995e+06	
1	2	Afghanistan	37171.921	2.412953	1.463960e+08	6.066142e+08	20257.50	743870.0	7.530102e+08	6.919553e+06	
2	3	Albanie	2882.740	-0.049546	6.496283e+07	5.674515e+07	42219.55	1226765.0	1.217080e+08	3.703744e+06	
2	3										





Analyse statistique

ACP et sélection

Conclusion

Merge des dataframes dispo_pays_volailles et df_pib :

	Code zone	Zo	ne	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Disponibilité en protéines (kg)	Disponibilité en protéines de volailles (kg)	Proportion protéine animale (%)	Proportion protéine volailles / prot animale(%)
0	1	Armér	nie	2951745.0	0.236146	4.881640e+07	5.278119e+07	34419.50	1092080.0	1.015976e+08	5.526995e+06	48.048780	11.322004
1	2	Afghanist	an :	37171921.0	2.412953	1.463960e+08	6.066142e+08	20257.50	743870.0	7.530102e+08	6.919553e+06	19.441441	4.726599
2	3	Albar	nie	2882740.0	-0.049546	6.496283e+07	5.674515e+07	42219.55	1226765.0	1.217080e+08	3.703744e+06	53.375983	5.701328
				173	pays			Zone	PIB par habitan	→ 18	5 pays		

		Zone	PIB par nabitant	
0	Afgh	anistan	2190.2	
1	Afrique	du Sud	12630.8	
2		Albanie	13601.3	

dispo_pays_ volailles



df pib

Jointure de type inner sur zone



- 165 pays

	Code zone	Zone	Population 2018	Différence population (%)	Dispo protéines animale (kg)	Dispo protéines végétale (kg)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Disponibilité en protéines (kg)	Disponibilité en protéines de volailles (kg)	Proportion protéine animale (%)	Proportion protéine volailles / prot animale(%)	PIB par habitant
0	1	Arménie	2951745.0	0.236146	4.881640e+07	5.278119e+07	34419.50	1092080.0	1.015976e+08	5.526995e+06	48.048780	11.322004	12715.0
1	2	Afghanistan	37171921.0	2.412953	1.463960e+08	6.066142e+08	20257.50	743870.0	7.530102e+08	6.919553e+06	19.441441	4.726599	2190.2
2	3	Albanie	2882740.0	-0.049546	6.496283e+07	5.674515e+07	42219.55	1226765.0	1.217080e+08	3.703744e+06	53.375983	5.701328	13601.3



Variables sélectionnées pour l'étude :

Zone	Population 2018	Différence population (%)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Proportion protéine animale (%)	Proportion protéine volailles/prot animale (%)	PIB par habitant
Arménie	2951745	0.236	34420	1092080	48.05	11.32	12715
Afghanistan	37171921	2.413	20258	743870	19.44	4.73	2190
Monténégro	627809	0.039	41723	1277135	61.46	8.70	20628
Soudan	41801533	2.421	26623	940240	27.61	2.68	4160

165 rows x 8 columns



Réduction et mise à l'échelle des variables avec MinMaxScaler :

```
array([[2.03090043e-03, 3.28505530e-01, 4.98558524e-01, ..., 6.03580373e-01, 2.02408443e-01, 8.99554037e-02], [2.60014017e-02, 7.28498221e-01, 1.25696713e-01, ..., 1.51081553e-01, 6.86647322e-02, 9.59802222e-03], [1.98256390e-03, 2.76009019e-01, 7.03920815e-01, ..., 6.87843821e-01, 8.84306037e-02, 9.67223491e-02], ..., [6.12940707e-03, 2.29186188e-01, 4.02364021e-01, ..., 6.32661202e-01, 1.66678935e-01, 1.25382802e-01], [4.03032976e-04, 2.92316201e-01, 6.90851432e-01, ..., 8.15645222e-01, 1.49188103e-01, 1.50376141e-01], [2.92443458e-02, 7.29996604e-01, 2.93292331e-01, ..., 2.80315820e-01, 2.71881277e-02, 2.46421261e-02]])
```

 $x_{std} = \frac{x - x_{minimum}}{x_{maximum} - x_{minimum}}$

Transformation des

variables [0; 1]

 $\int x_{scaled} = x_{std} \times (x_{maximum} - x_{minimum}) + x_{minimum}$

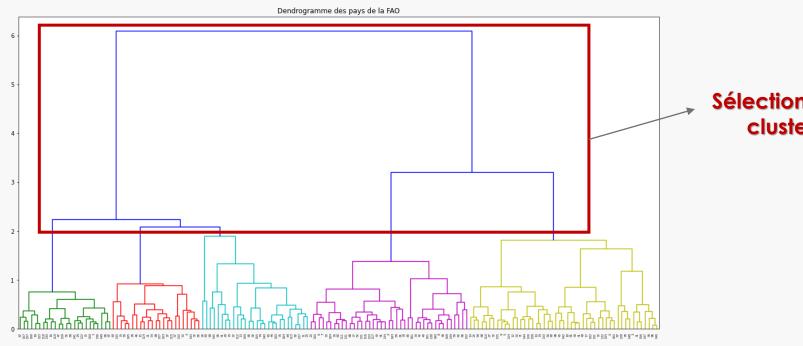


Analyse Données Clustering statistique

ACP et sélection

Conclusion

Traçage du dendrogramme :



Sélection de 5 clusters



Données

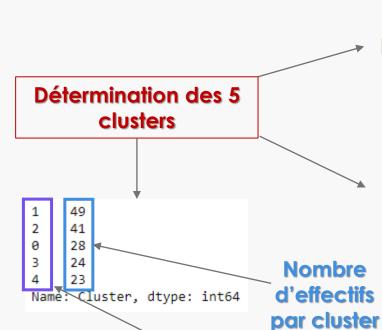
Clustering

Analyse statistique

ACP et sélection

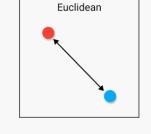
Conclusion

Clustering agglomératif :

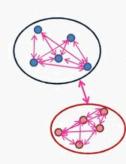


Clusters

Distance euclidienne



Méthode Ward:
analyse les distances qui
minimisent la variance
dans les clusters et
maximisent la variance
entre les clusters





Données Clustering Analyse ACP et sélection Conclusion

Visualisation des clusters :

	Population 2018	Différence population (%)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Proportion protéine animale (%)	Proportion protéine volailles / prot animale(%)	PIB par habitant
Cluster		(/			V		
0	1.308126e+08	1.707972	27175.292857	9.754234e+05	38.026025	7.886916	9530.417857
1	2.671928e+07	0.730943	31403.631633	1.085413e+06	47.431958	27.097162	19840.753061
2	2.852243e+07	0.337384	39154.084146	1.241881e+06	60.020154	15.144094	46761.160976
3	3.147140e+07	2.766663	21587.164583	8.554535e+05	20.078773	7.207911	2571.095833
4	1.563371e+07	1.984027	22753.147826	9.270683e+05	33.342290	26.227630	7051.413043

Moyenne de chaque variable en fonction du cluster (0 à 4)



Données

Clustering

Analyse statistique

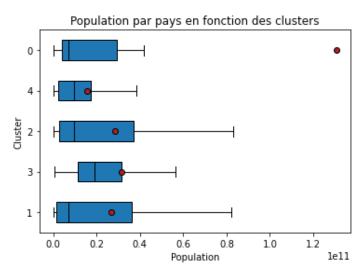
ACP et sélection

Conclusion

Analyse des caractéristiques clusters :



Population des clusters



Test d'Anova Welch	Test de shapiro
p-value =0.18	p-value = 1e-25

Anova Welch

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value > 0.05: Les moyennes de population sont égales dans les clusters

Test de Shapiro

H0: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value < 0.05: La population ne suit pas une distribution normale



Données

Clustering

Analyse statistique

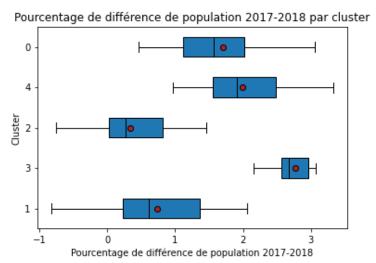
ACP et sélection

Conclusion

Analyse des caractéristiques clusters :



Différence de population



Test d'Anova Welch	Test de shapiro
p-value =4e-28	p-value = 0.27

Anova Welch

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05: Il y a une différence significative des moyennes de la différence de population dans les clusters

Test de Shapiro

H0: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value > 0.05: La différence de population suit une distribution normale



Clustering

Analyse statistique

ACP et sélection

Conclusion

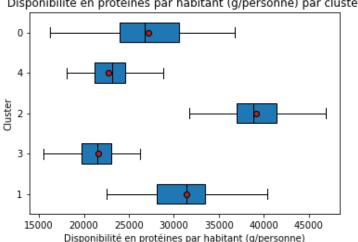
Analyse des caractéristiques clusters :



Disponibilité en protéine par habitant

Anova Welch

Disponibilité en protéines par habitant (g/personne) par cluster



Test d'Anova Welch	Test de shapiro		
p-value =2e-29	p-value = 0.02		

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05: If y a une différence significative des moyennes de disponibilité en protéine dans les clusters

Test de Shapiro

HO: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value < 0.05): La disponibilité en protéine ne suit pas une distribution normale



Données

Clustering

Analyse statistique

ACP et sélection

Conclusion

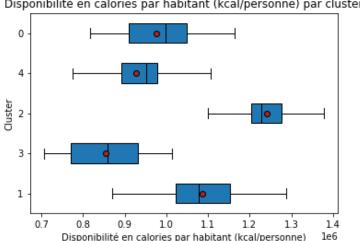
Analyse des caractéristiques clusters :



Disponibilité en calories par habitant

Anova Welch

Disponibilité en calories par habitant (kcal/personne) par cluster



Test d'Anova Welch	Test de shapiro
p-value =5e-29	p-value = 0.09

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05: If y a une différence significative des moyennes de disponibilité en calories dans les clusters

Test de Shapiro

HO: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value > 0.05): La différence de population suit une distribution normale



Clustering

Analyse statistique

ACP et sélection

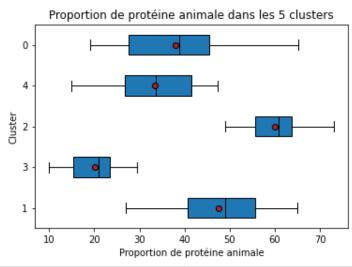
Conclusion

Analyse des caractéristiques clusters :



Proportion de protéine d'origine animale

Anova Welch



Test d'Anova Welch Test de shapiro
p-value = 5e-38 p-value = 3e-5

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05: Il y a une différence significative des moyennes de proportion de protéine animale dans les clusters

Test de Shapiro

HO: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value < 0.05: La proportion de protéine animale ne suit pas une distribution normale



Clustering

Analyse statistique

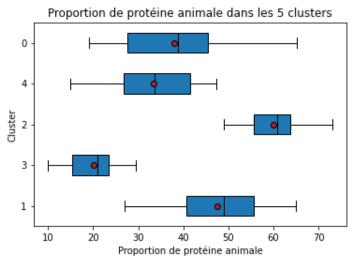
ACP et sélection

Conclusion

Analyse des caractéristiques clusters :



Proportion de viande de volailles



Test d'Anova Welch	Test de shapiro	
p-value =3e-22	p-value = 9e-6	

Anova Welch

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de proportion de protéine de volailles dans les clusters

Test de Shapiro

H0: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value < 0.05: La proportion de protéine de volailles ne suit pas une distribution normale



Données

Clustering

Analyse statistique

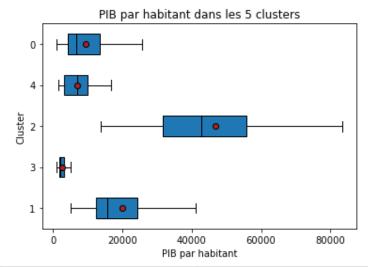
ACP et sélection

Conclusion

Analyse des caractéristiques clusters :



PIB par habitant



Test d'Anova Welch	Test de shapiro		
p-value =8e-23	p-value = 9e-4		

Anova Welch

HO: Les moyennes des échantillons sont égales

Ha: Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de PIB par dans les clusters

Test de Shapiro

H0: L'échantillon suit une distribution normale

Ha: L'échantillon ne suit pas une loi normale

p value < 0.05: Le PIB par habitant ne suit pas une distribution normale



Analyse statistique

p-value

ACP et sélection

Conclusion

Analyse des caractéristiques clusters :

Test de corrélation pearson

	Dispo_prot	Prop_volailles	
Dispo_cal	0.88 / 1e-57	0.11 / 0.13	
Diff_pop	-0.55 / 1e-14	-0.21 / 0.36	
Prop_anim	0.71 / 1e-26	0.26 / 0,0005	
PIB_habitant	0.71 / 1e-26	0.07 / 0.36	
Prop_volailles	0.05 / 0.46	х	

Le test de Pearson mesure une dépendance linéaire entre 2 variables : si la p-value est inférieure à 0.05 la corrélation est significative

Coefficient de corrélation



Données

Clustering

Analyse statistique

ACP et sélection

Conclusion

Test de comparaison des variables gaussiennes entre le cluster 1 et 4 :

→

Différence population (%)

→

Disponibilité en calories par habitant

Kolmogorov-smirnov

Kolmogorov-smirnov

HO: Les échantillons suivent une même distribution

Ha: Les échantillons ne suivent pas une même

distribution

p value < 0.05): Les clusters ne suivent pas une

même distribution

HO: Les échantillons suivent une même distribution Ha: Les échantillons ne suivent pas une même distribution

p value < 0.05 : Le PIB par habitant ne suit pas une distribution normale

Test kolmogorov-smirnov:

p-value = 2e-7

Test kolmogorov-smirnov : p-value = 8e-9

26

Centrage et réduction des données

```
array([[-9.48370000e-01, 5.84999107e-01, 2.54158346e-01,
         3.17084084e-01, -5.88992678e-01, -3.67565192e-01],
       [ 1.04897814e+00, -1.28042912e+00, -1.85436201e+00,
       -1.48711990e+00, -1.18380354e+00, -8.65401495e-01],
      [-1.21050931e+00, 1.61242697e+00, 1.06971811e+00,
        6.53059432e-01, -1.09589694e+00, -3.25642085e-01],
      [ 6.95420663e-01, 4.65765551e-01, 9.81310543e-01,
       -1.01532191e+00, -8.36638141e-01, -4.26005899e-01],
      [ 1.89080058e+00, -1.39822033e+00, -1.09184679e+00,
       -7.74254970e-01, 6.01712794e-01, -6.41037428e-01],
      [-3.38123603e-01, -1.06843990e-01, -9.61445632e-01,
        1.37025830e+00, 1.68255516e+00, 2.98032340e-02],
      [-2.79568774e-01, 1.18357079e+00, 9.52578086e-01,
        1.36622192e+00, 2.41963366e-01, 1.06909018e-01],
      [ 5.13124547e-03, 1.14030055e+00, 1.13602378e+00,
        1.49921884e+00, 5.46981998e-01, 1.37600656e+00],
```

Transformation des variables [0 ; 1]

$$x_{scaled} = \frac{x - x_{moyenne}}{\sigma_{écart_type}}$$

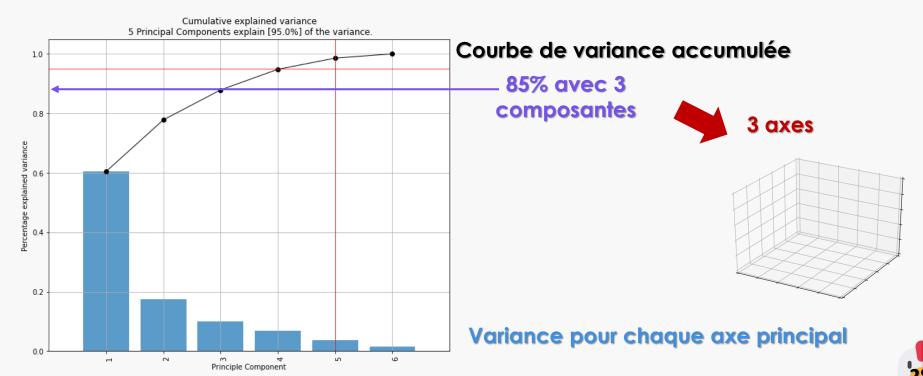


Analyse statistique

ACP et sélection

Conclusion

Détermination du nombre de composantes :

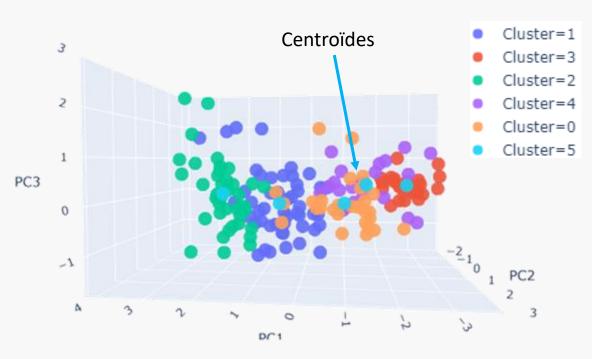


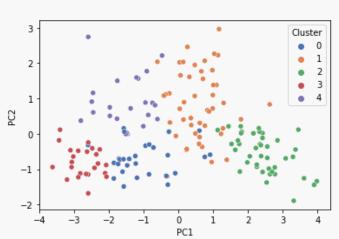
Analyse statistique

ACP et sélection

Conclusion

Analyse en composantes principales :

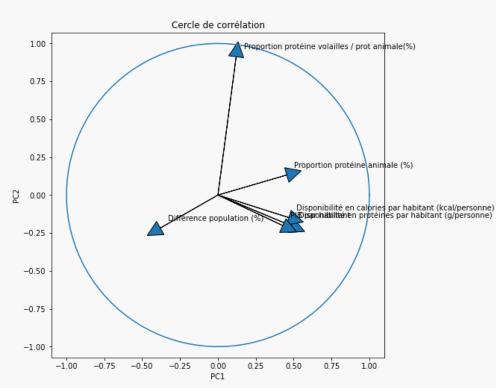






Données Clustering Analyse ACP et sélection Conclusion

Cercle de corrélation :



Corrélation positive entre la disponibilité en calories, la disponibilité en protéine et le PIB par habitant

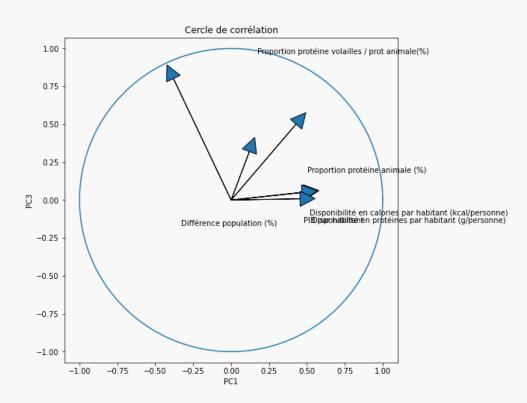


Analyse statistique

ACP et sélection

Conclusion

Cercle de corrélation :



Corrélation positive entre la proportion de protéine de volailles et la proportion de protéine animale



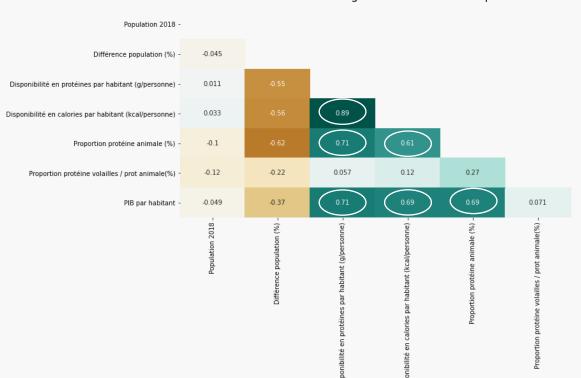
Analyse statistique

ACP et sélection

Conclusion

Heatmap de corrélation :

Triangle Correlation Heatmap





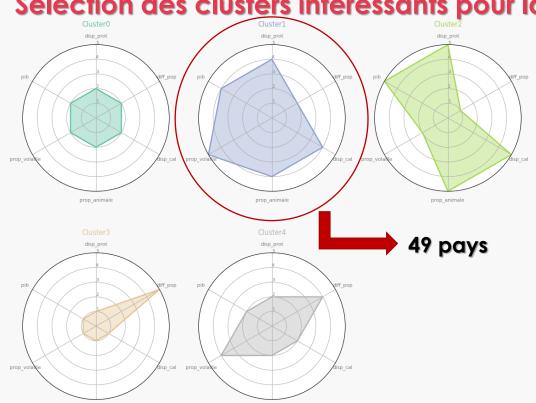


Analyse statistique

ACP et sélection

Conclusion

Sélection des clusters intéressants pour la commercialisation de poulet :



Cluster 1 possède les pays avec le plus de proportion de protéines de volailles



Analyse statistique

ACP et sélection

Conclusion

Sélection des pays intéressants pour la commercialisation de poulet :

Zone	Population 2018	Différence population (%)	Disponibilité en protéines par habitant (g/personne)	Disponibilité en calories par habitant (kcal/personne)	Proportion protéine animale (%)	Proportion protéine volailles / prot animale(%)	PIB par habitant
Brésil	2.094693e+11	0.786927	34915.90	1204500.0	56.199038	28.571429	14596.2
Colombie	4.966105e+10	1.535906	26608.50	1135880.0	51.193416	27.706324	14455.6
Israël	8.381516e+09	1.669948	46289.30	1287355.0	58.894496	37.407953	39543.2
Liban	6.859408e+09	0.587077	25203.25	1040615.0	32.541636	28.615932	15612.0
Malaisie	3.152803e+10	1.361170	28261.95	1037695.0	55.779414	29.937486	27536.9
Arabie saoudite	3.370276e+10	1.817388	31718.50	1207055.0	37.675489	38.485034	47596.7
Afrique du Sud	5.779252e+10	1.373032	30795.05	1056675.0	42.716605	42.286349	12630.8
Trinité-et- Tobago	1.389843e+09	0.417901	31240.35	1089160.0	51.793434	43.469434	26272.9
Émirats arabes unis	9.630959e+09	1.515262	40193.80	1208880.0	45.813658	28.800793	66968.3

10 pays avec une grande proportion en protéine animale et de volailles ainsi qu'une population élevée



Analyse statistique

ACP et sélection

Conclusion

Carte des pays sélectionnées pour la commercialisation du poulet :

Carte du monde représentant les pays sélectionnés pour la commercialisation du poulet





- ☐ Détermination des **pays propices** à la commercialisation du poulet
- Clustering des pays possédant des caractéristiques proches
- Pays avec les proportions en protéines de volailles et animales les plus élevées ainsi qu'une grande population
- ☐ 10 Pays sélectionnés :

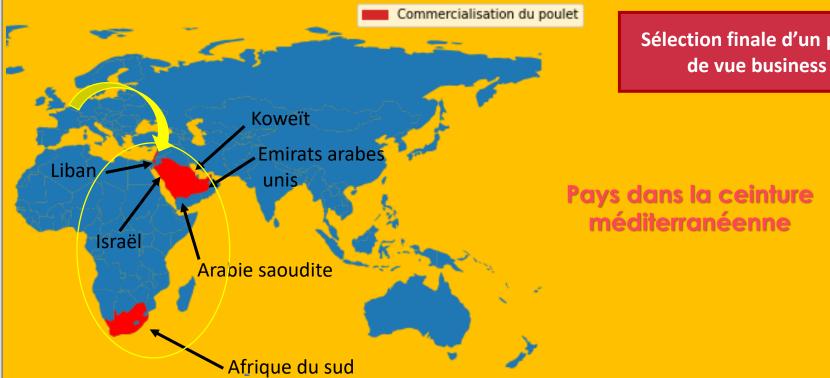
Barrière linguistique

Pays proches pour l'exportation

Brésil	Colombie	Israël	Koweït	Liban
Malaisie	Trinité-et- Tobago	Afrique du sud	Arabie saoudite	Emirats arabes unis

Différents fuseaux horaires







Merci pour votre attention

