1)Je vais vous présenter le projet 5 produiser une étude de marché

3) La problématique c’est de savoir quels pays sont les plus propices pour exporter nos poulets ?

4) ***Objectif:*** ***Récupérer les données qui nous intéresse et les regrouper le plus clairement possible***

5) J’ai importé les données de la FAO (un fichier par variable qui m’intéresse, donc un fichier pop2008, un pop2018 etc..)

Puis je les ai modifié, j’ai gardé les colonnes qui m’intéressait « Zone »: nom des pays et « Valeur » : La valeur de la variable étudié.

J’ai modifié le nom de la colonne valeur par le nom de la variable étudié (j’ai remplacé Valeur par pop2008 pour le premier tableau etc..)

J’ai affiché les 4 premiers tableaux obtenus

6) J’ai joint mes 6 tableaux pour faire mon dataframe principale.

J’ai regardé si il y avait des valeurs non renseignées. J’en ai eu 2.

Je les ai remplacé par des valeurs trouvées sur internet qui semblaient cohérentes avec le reste des données. (wikipedia)

7) J’ai ensuite effectué des calculs pour obtenir les variables souhaité pour l’analyse,

La quantité de protéine disponible par habitant, les kcal disponible par habitant, l’évolution

de la population en pourcentage entre 2008 et 2018,

Le pourcentage de protéine animal dans l’ensemble des protéines disponible et j’ai ajouté

le PIB pour mieux caractériser les groupes par la suite.

8) ***Objectif:*** ***Faire 5 groupes les plus cohérents possibles selon les critères choisis***

9) J’ai effectué une classification hiérarchique ascendante pour former 5 groupe (Je me base sur les 3 axes principaux de mon acp)

J’ai utilisé la méthode de Ward.

Cette méthode cherche à minimiser l’inertie intra-classe et à maximiser l’inertie inter-classe afin d’obtenir des classes les plus homogènes possibles.

(Cette méthode est souvent incorrectement présentée comme une méthode de minimisation de la variance alors qu’au sens strict Ward vise l’augmentation minimum de la somme des carrés)

J’ai obtenu un dendrogramme je vais maintenant vous présenter chacun des groupe obtenu. Je vais tenter une première approche sans calcul pour voir si il y a une cohérence dans ces groupes si les membres de chaque groupes sont proche et si les groupes sont bien différents.

10) Voici la liste des pays du premier cluster

Dans se groupe on retrouve surtout des pays d’afrique centrale et des pays ravagé par la guerre (iraq, afghanistan, tadjikistan)

11) Dans ce groupe il y a énormément de diversité (beaucoup de pays d’Asie, les pays du magrheb, quelque pays d’europe de l’est et d’amérique latine) difficil à définir sans chiffre

12) Les pays sont casiment tous des pays d’europe. Il y a aussi certain pays très occidentalisé (canada, australie, israel, états unis….)

13) Ces pays sont pour la plupart en transition démographique et/ou économique

14) Ce cluster est très particulier il ne rassemble que 4 pays. Ce sont des pays assez proche géographiquement, mais je pense mieu en saisir la cohérence grace au chiffre

16) Je fais une analyse en composante principale et j’observe les pays dans mon premier plan qui représente 86.7% de l’inertie totale.

Au vue de la grande inertie de ce plan nous n’observerons pas les autres plans.

Pour simplifier de façon drastique le schéma on peut dire que :

L’Axe des x va majoritairement expliquer la richesse du régime alimentaire

(plus on va vers la droite plus le régime alimentaire est riche en calorie et protéine, particulièrement en protéine animale)

L’axe des y nous donnera un renseignement sur l’évolution de la démographie. (plus on est haut plus l’augmentation de pop a été forte)

17) À priori les groupes 5 et 3 sont les plus intéressants

18) Je vais maintenant m’intéresser aux centroides de mes groupes.

Le groupe 3 domine largement au niveau du régime alimentaire et le groupe 5 au niveau la de la démographie

18)Je souhaite caractériser au mieux mes clusters:

Je décide donc de calculer mes centroide.

Pour avoir une approche plus complète je vais m’intéresser aussi au pib moyen des clusters.

J’analyse mes cluster sur 3 plans la démographie, le régime alimentaire et l’économie.

Cluster 1 régime alimentaire très faible en protéine et en calorie, pourcentage de nourriture animal très faible, pib extrêmement faible, très forte augmentation de population

Cluster 2 pib assez faible, régime alimentaire équilibré

Cluster 3 régime alimentaire très riche en protéine et calorie, pib très élevé et évolution de pop assez faible.

Cluster 4 pays similaire au cluster 1 mais avec un pib un peu plus élevé un prct de prot animal un peu plus élevé et une pop\_diff plus faible

Cluster 5 régime alimentaire riche en protéine animal, un pib moyen, augmentation de pop extrêmement élevé

19)Grâce à ces caratérisation je vais donné des nom à mes cluster j’opte pour :

21) Malgré l’augmentation faible de la population de ce groupe durant les 10 dernières années

Le groupe « pays occidentalisé » est clairement plus intéressant ! Son régime alimentaire bien plus riche en calorie et en protéine animal font de de lui le meilleur candidat.

22) Je décide de classer les pays de mon groupe selon 2 critères qui me semble pertinents

Le pib et la démographie.

Le groupe 3 a été sélectionner pour le régime alimentaire riche des pays qui le compose mais l’augmentation de population a était négligé et le pib ne faisait pas partie des variables prise en compte lors de la création des groupes.

Nous décidons donc maintenant d’en tenir réellement compte.

Si la stratégie est d’exporter nos produits il faut que le pays est un pib par habitant élevé. Nos poulets seront d’autant mieux vendu si leur prix sont moins importants que les poulets locaux.

Ne sachant pas quelle importance donner à mes critères je décide de faire une acp

Plus les points seront à droite plus ils seront attractifs, l’axe y nous permet juste de savoir si cette attractivité est plutôt lié au pib ou a l’augmentation de pop.

23) Des pays se détachent très largement !

24) Voici les 15 pays les plus propices à notre développement.

Je me suis arrêté à 15 car il y a une différence conséquente entre le 15em et le 16em pays

25) ***Objectif :****Confirmer que nos groupes soient significativement différents*

26)Pour cela je vais effectuer des tests d’adéquations

1. Sélection des deux clusters les plus proches graphiquement selon une variable distribué normalement

2. Test d’adéquation sur les variances : test de Fisher

3. Si les variances sont similaires, on peut tester la moyenne dans de bonnes conditions avec un test de Student

26) Pour savoir si une de mes variable est distribué normalement

J’ai effectué un shapiro test sur toute mes variable + le pib.

Je considère que pour une p-value supérieur à 0.05 mon échantillon est distribué normalement selon ma variable.

Aucune variable ne passe le test.

J’ai aussi effectué le test de kolgomorov et je n’ai pas obtenu de meilleurs résultats..

D’après les tests et les courbe de densitées, la variable

kcal (quantité disponible de calorie par habitant) est à priori la meilleur variable.

27) Je regarde la dispersion et les moyennes de mes 5 groupes par rapport à la variable kcal

Je remarque que deux de mes groupes sont très proches. C’est eux que je vais sélectionner pour mes tests (si ils passent le test à priori les autres groupes aussi)

28) La variable kcal est distribué assez normalement dans mes 2 groupes je peux donc effectuer mon test de Fisher d’égalité de deux variance.

(les courbes semblent assez différentes.)

30) On test d’abord la variance à l’aide d’un test de fisher

**H0 :** Les variances sont similaires

**H1:** Les variances sont différentes

Je choisis un seuil élevé de 10% pour plusieurs raisons. Dabord parceque mes deux groupes sont très différents pour 3 autres variables pop\_diff, prot et PIB.

Aussi car la normalité de la variable kcal est loin d’être certaine.

Et aussi parceque les courbes me semble bien différentes.

J’obtient une p-value de 0.09905 donc inférieur à 0.1. Je suis en dessous du seuil de 10%

Je rejette donc l’hypothèse nul. Les variances de mes 2 groupe sont significativement différents

Donc l’ensemble de mes groupes sont significativement différents.

30)Conclusion

Les tests d’adéquations valident mes groupes et donc mes choix.

Les 15 pays les plus propices pour notre expansion à l’international sont donc

bien ceux de la liste.

On peut ajouter de nombreux autres critères si l’on souhaite affiner la sélection (distance qui nous sépare du pays, % de poulet dans l’alimentation… etc)

« « « « 30) Ce projet m’a permis de me familiariser avec les algorithmes de tri et les tests statistiques.

J’ai trouvé certains points de la démarche trop guidées.

J’aurais aimé utilisé la fonction hcpc qui combine cah et kmeans. J’aurai aussi aimé pouvoir choisir mon nombre de groupe. » » »