



Étude de Marché



La poule qui chante

Contexte

- La Poule Qui Chante – une entreprise française d'agroalimentaire.
- Nous voulons développer à l'international – mais où ?
- Champs des possibles est large – avec cette étude nous pouvons choisir le meilleur pays.

Données pour l'étude

- *Population*
- *Disponibilité alimentaire*

Politique
Stabilité Politique

Economique
Produit intérieur brut par habitant

Socioculturel
Poulet disponible par habitant

Technologique
Crédit à l'agriculture

Ecologique
Émissions provenant des élevages de poulets

Légal
Investissement gouvernemental dans l'agriculture

<https://www.fao.org/faostat/en/?#data>



Préparation & Nettoyage

Étapes générales

- Importer les librairies python
- Importer les fichiers csv
- Pour chaque fichier ;
 - Vérifier la tête, la forme et les types de données
 - Vérifier si il y a des valeurs 'nulles'
 - Supprimer les colonnes qui ne sont pas utile pour l'analyse
 - Vérifier que « Zone » peut être utiliser comme clé primaire.
- Fusionner les fichiers





Fichier final fusionné

- Trouver des valeurs nulles :
 - PIB (1) – Les données de Taiwan étaient manquantes. J'ai utilisé des données de FMI
 - Stabilité politique (4) – Données de la FMI pour la Chine et Taiwan. Stabilité politique de France pour la Nouvelle Calédonie et Polynésie Française.
 - Investissements des gouvernements (29) et crédits à l'agriculture (62) – La FAO a expliqué que les taux de réponse pour ces données sont relativement faibles. J'ai remplacé les valeurs nulles avec les valeurs moyennes de chaque colonne.
 - Emissions (4) - le Djibouti et les Maldives ne produisent pas de poulet (pas d'émissions), Le FAO a les données pour Islande pour 2013 et 2019, and données pour la Lettonie jusqu'à 2016. J'ai respectivement interpolé et extrapolé les données.



Analyse exploratoire du fichier fusionné

- Vérifier les types de données dans chaque colonne
- Utiliser 'describe' afin de voir les min, max, quartiles et écart type de chaque colonne.
- Créer un boxplot et un histogramme pour chaque colonne
- Trouver les outliers pour chaque colonne.





Les Résultats



Classification Ascendante Hiérarchique

Démarche

- Normalisation des données
- Créer des clusters à l'aide de la méthode Ward
- Créer un dendrogramme avec ces clusters
- Déterminer le nombre optimal de clusters à l'aide de l'indice Davies-Bouldin
- Créer un tableau de contingence et une heatmap
- Créer une liste des clusters avec les pays dans chaque cluster
- Effectuer une analyse plus approfondie des clusters – fonction « describe() », boîtes à moustaches et histogrammes



Classification Ascendante Hiérarchique

Cluster 1

Politique
Stabilité politique - faible
Economique
PIB par habitant - faible
Socioculturel
Proportion de poulet mangé - faible
Technologique
Crédit à l'agriculture - faible
Ecologique
Emissions - élevé
Légal
Investissement gouvernemental - faible

Cluster 2

Politique
Stabilité politique – moyen
Economique
PIB par habitant – moyen
Socioculturel
Proportion de poulet mangé - élevé
Technologique
Crédit à l'agriculture - moyen
Ecologique
Emissions - moyen
Légal
Investissement gouvernemental - moyen

Cluster 3

Politique
Stabilité politique - élevé
Economique
PIB par habitant - élevé
Socioculturel
Proportion de poulet mangé - moyen
Technologique
Crédit à l'agriculture - élevé
Ecologique
Emissions - faible
Légal
Investissement gouvernemental - élevé





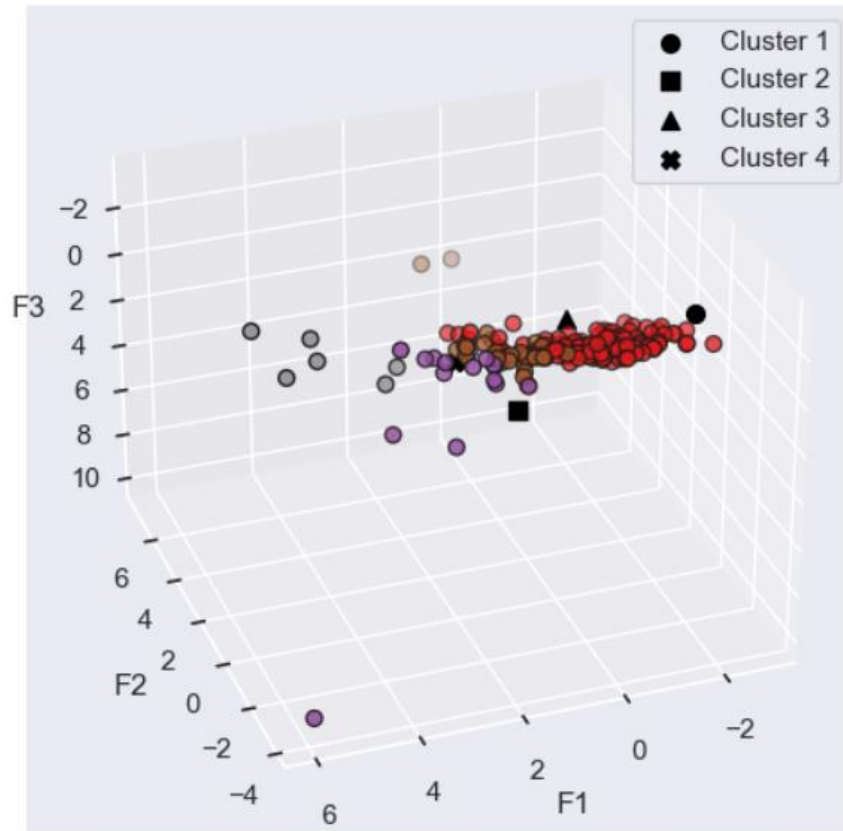
Analyse de cluster K-means

Démarche

- Déterminer le nombre optimal de clusters (4) – utilisez la méthode du coude
- Utiliser le nombre optimal de clusters pour instancier les k-moyennes, puis entraînez-les
- Ajouter les centroïdes de chaque cluster de k-moyennes à une variable appelée « centroïdes »
- Mettre à l'échelle les valeurs des colonnes et entraînez l'ACP avec les données mises à l'échelle et 4 composants
- Visualiser les clusters et leurs centroïdes
- Créer une liste des clusters avec les pays dans chaque cluster
- Effectuer une analyse plus approfondie des clusters – fonction « describe() », boîtes à moustaches et histogrammes



Analyse de cluster K-means



- Le centroïde de cluster 1 est relativement éloigné du centroïde de cluster 2 et du centroïde de cluster 4.
- En revanche, le centroïde de cluster 3 est proche de 1, 2 et 4.
- Les centroïdes de clusters 2 et 3 sont les clusters les plus similaires.



Analyse de cluster K-means

Pays pauvres

Politique
Stabilité politique - faible
Economique
PIB par habitant - faible
Socioculturel
Poulet disponible - faible
Technologique
Crédit à l'agriculture - faible
Ecologique
Emissions - élevé
Légal
Investissement gouvernemental - faible

Poulet, PIB élevé

Politique
Stabilité politique – moyen
Economique
PIB par habitant – élevé (2 ^e)
Socioculturel
Poulet disponible - élevé
Technologique
Crédit à l'agriculture - élevé
Ecologique
Emissions - faible
Légal
Investissement gouvernemental - moyen

Poulet, faible PIB

Politique
Stabilité politique - moyen
Economique
PIB par habitant - moyen
Socioculturel
Poulet disponible - élevé
Technologique
Crédit à l'agriculture - moyen
Ecologique
Emissions - moyen
Légal
Investissement gouvernemental - moyen

Pays riches

Politique
Stabilité politique - élevé
Economique
PIB par habitant – élevé (1 ^{er})
Socioculturel
Poulet disponible - moyen
Technologique
Crédit à l'agriculture - moyen
Ecologique
Emissions - faible
Légal
Investissement gouvernemental - élevé



Pays à examiner

CAH

- Finlande
- Irlande
- Islande
- Luxembourg
- Norvège
- Nouvelle-Zélande
- Suisse

K-means

- Allemagne
- Australie
- Autriche
- Belgique
- Belize
- Canada
- Chine - RAS de Hong-Kong
- Danemark
- Finlande
- Israël
- Japon
- Nouvelle-Zélande
- Pays-Bas
- Royaume-Uni
- Suède
- Émirats arabes unis
- États-Unis d'Amérique





Analyse en composantes principales

Démarche

- Séparer les valeurs quantitatives de le dataframe d'origine.
- Entraîner le scaler avec ces valeurs et transformer les données.
- Créer un ACP et entraîner-le avec les données mises à l'échelle
- Calculer la variance des données pour comprendre la contribution de chaque fonctionnalité et déterminer le nombre optimal de composants.
- Créer un heatmap
- Tracer un cercle de corrélation et projeter les différents pays sur les plans factoriels.
- Analyser les pays à partir de ces graphiques.



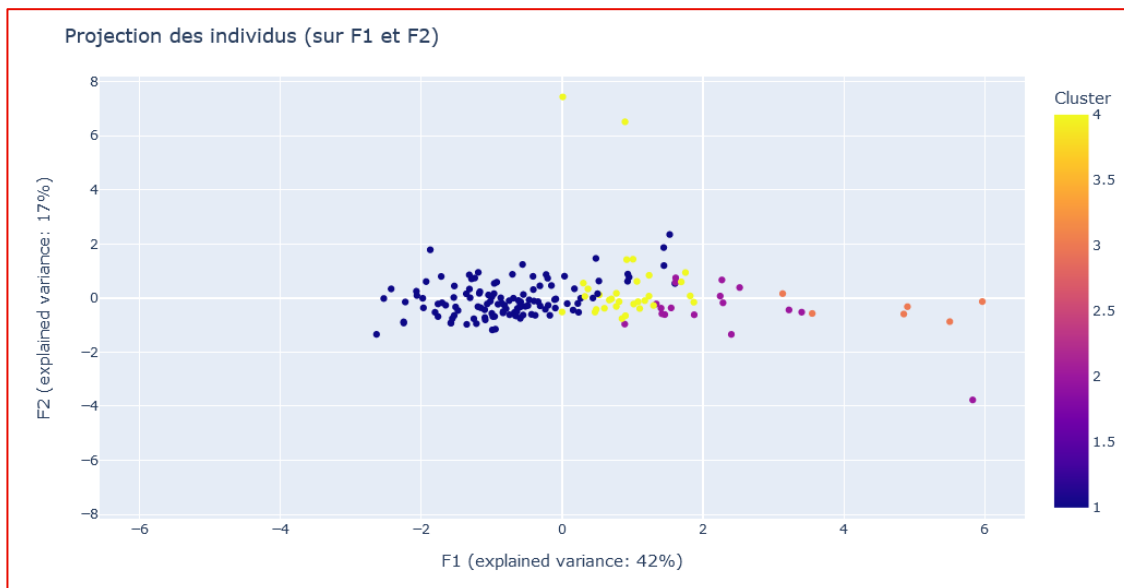
ACP – Composantes Principales

	F1	F2	F3	F4
Viande de volaille - Disponibilité alimentaire en quantité (kg/personne/an)	0.37	0.44	0.13	-0.65
PIB (US\$) par habitant	0.55	-0.07	-0.15	0.16
Stabilité Politique	0.50	0.17	0.05	-0.14
Crédit à l'Agriculture, Sylviculture et Pêche par habitant - (US\$)	0.27	-0.33	0.86	0.23
Viande, poulet – Intensité des émissions (Kilogrammes d'équivalent CO2 par kilogramme)	-0.16	0.82	0.26	0.47
Investissements de l'administration centrale dans l'agriculture, sylviculture et pêche par habitant (US\$)	0.46	0.02	-0.40	0.51

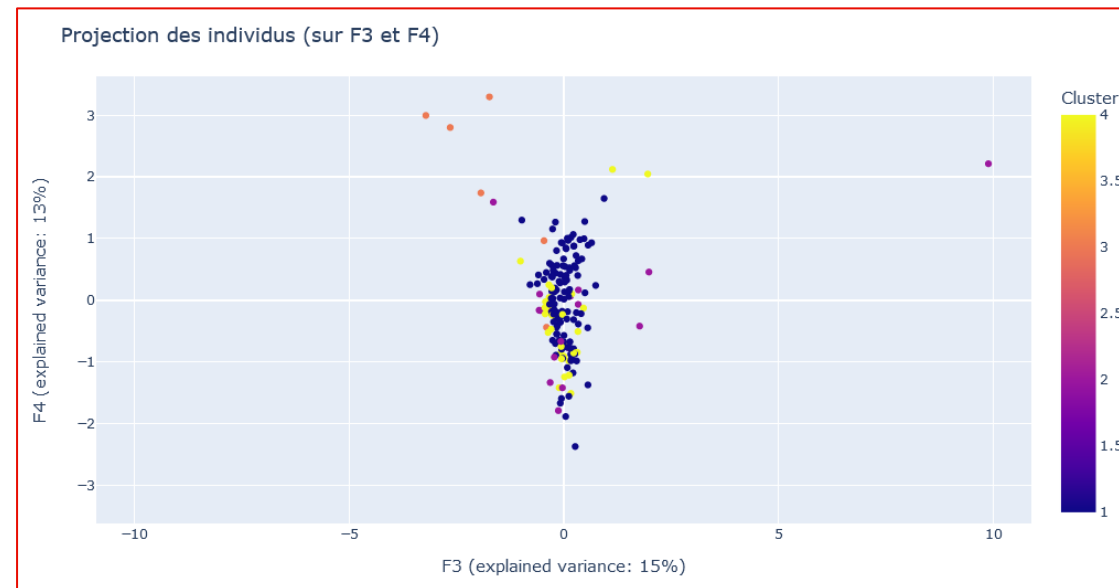
- F1 – Prospérité économique
- F2 – Quantité d'émissions
- F3 – Crédit à l'agriculture
- F4 – Besoin d'investissement dans l'agriculture



Analyse en composantes – Visualisations – K-means



- F1 – Prospérité économique
- F2 – Quantité d'émissions



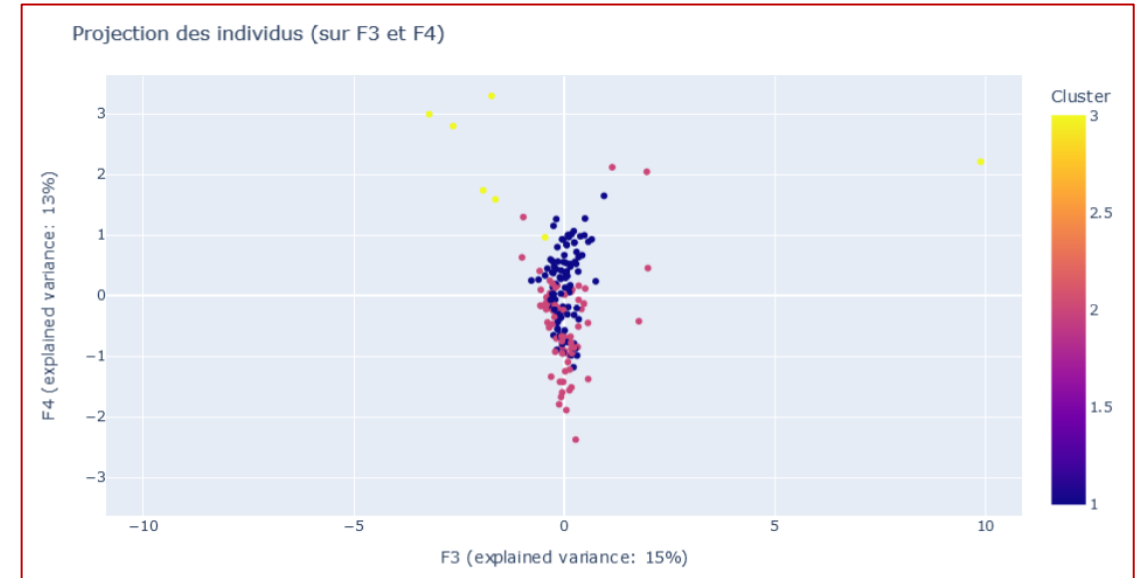
- F3 – Crédit pour l'agriculture
- F4 – Besoin d'investissement dans l'agriculture



Analyse en composantes – Visualisations - CAH



- F1 – Prospérité économique
- F2 – Quantité d'émissions



- F3 – Crédit pour l'agriculture
- F4 – Besoin d'investissement dans l'agriculture



Zones pour le développement

- L' Australasie (Australie et Nouvelle Zélande)
 - Une grande population (24.5 millions en Australie)
 - Beaucoup de poulet dispo
 - Loin de la France – des régulations plus complexe
- L'Union Européenne
 - Surtout Pays-Bas et Irlande
 - Plus simple et moins chère pour les exportations des produits.
 - Moins poulet dispo pourrait suggérer que la demande est plus faible qu'en Australasie

Europe



Australasie