



Produisez une étude de marché

[MISSION](#)[COURS](#)[ÉVALUATION](#)

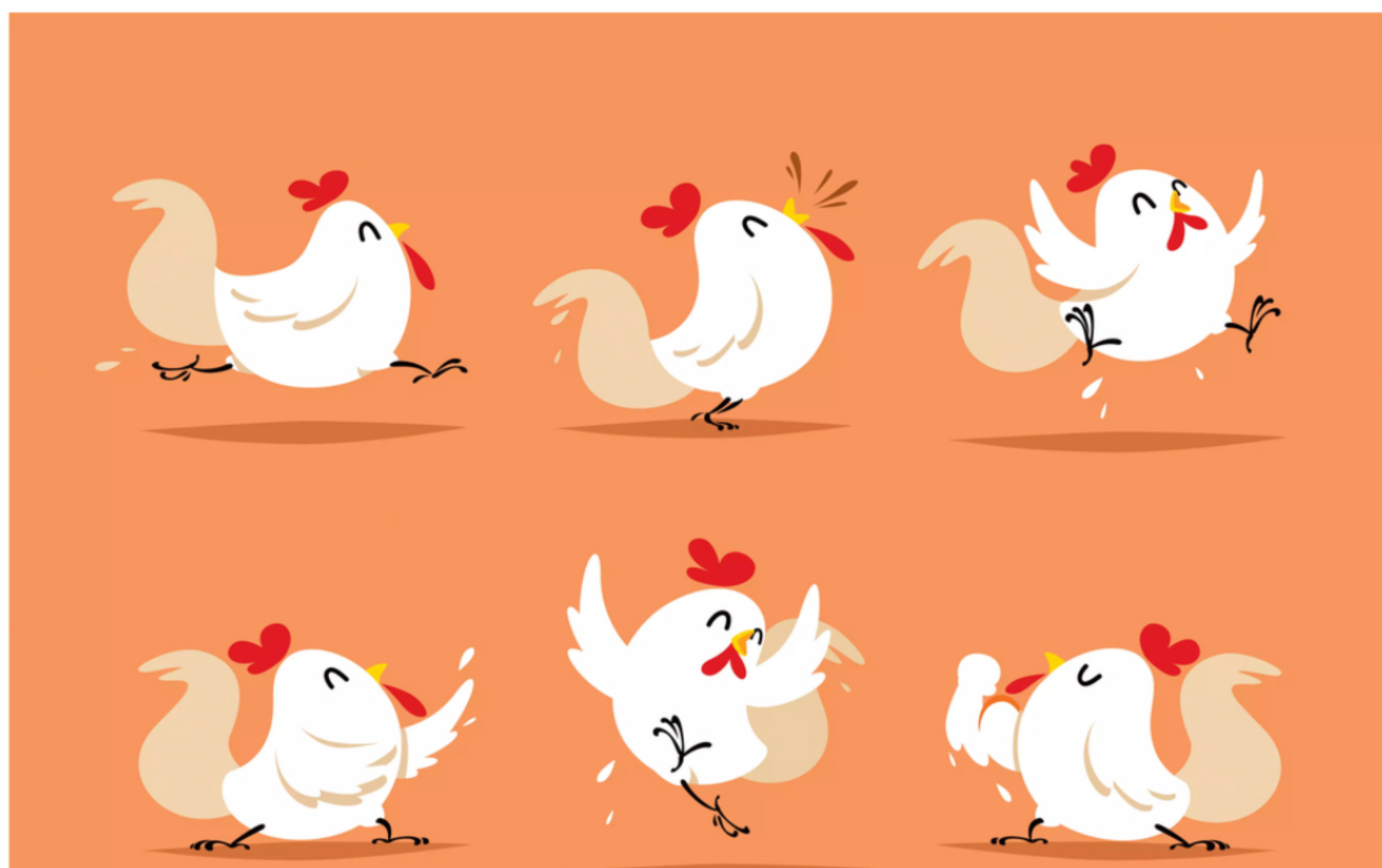
Mis à jour le lundi 1 avril 2019

Prérequis

Pour effectuer ce projet, vous devrez maîtriser de la manipulation de données en **Python** ou **R**, appliquer ces langages à la **statistique descriptive** ainsi qu'à la **classification automatique**.

Mise en situation

Votre entreprise d'agroalimentaire souhaite se développer à l'international. Elle est spécialisée dans...



... le poulet !

L'international, oui, mais pour l'instant, le champ des possibles est bien large : aucun pays particulier ni aucun continent n'est pour le moment choisi. Tous les pays sont envisageables !

Votre objectif sera d'aider à cibler plus particulièrement certains pays, dans le but d'approfondir ensuite l'étude de marché. Plus particulièrement, l'idéal serait de produire des "groupes" de pays, plus ou moins gros, dont on connaît les caractéristiques.

Dans un premier temps, la stratégie est plutôt d'exporter les produits plutôt que de produire sur place, c'est-à-dire dans le(s) nouveau(x) pays ciblé(s).

Les données

Vous vous souvenez de la FAO, dans l'un de vos précédents projets ? Allez, [on y retourne](#) ! Vous connaissez déjà l'interface du site, à vous de retrouver les données qui vous seront utiles pour le projet.

Votre mission

Pour identifier les pays propices à une insertion dans le marché du poulet, il vous a été demandé de cibler les pays. Il vous faudra également étudier les régimes alimentaires de chaque pays, notamment en termes de protéines d'origine animale et en termes de calories.

Construisez votre échantillon contenant l'ensemble des pays disponibles, chacun caractérisé par ces variables :

- Différence de population entre une année antérieure (au choix) et l'année courante, exprimée en pourcentage.
- Proportion de protéines d'origine animale par rapport à la quantité totale de protéines dans la disponibilité alimentaire du pays

disponibilité alimentaire du pays.

- Disponibilité alimentaire en protéines par habitant.
- Disponibilité alimentaire en calories par habitant.

Construisez un dendrogramme contenant l'ensemble des pays étudiés, puis coupez-le afin d'obtenir 5 groupes.

Caractérisez chacun de ces groupes selon les variables cités précédemment, et facultativement selon d'autres variables que vous jugerez pertinentes (ex : le PIB par habitant). Vous pouvez le faire en calculant la position des centroïdes de chacun des groupes, puis en les commentant et en les critiquant au vue de vos objectifs.

Donnez une *courte* liste de pays à cibler, en présentant leurs caractéristiques. Un découpage plus précis qu'en 5 groupes peut si besoin être effectué pour cibler un nombre raisonnable de pays.

Visualisez vos partitions dans le premier plan factoriel obtenu par ACP.

Dans votre partition, vous avez obtenu des groupes distincts. Vérifiez donc qu'ils diffèrent réellement. Pour cela, réalisez les tests statistiques suivants :

- **un test d'adéquation** : parmi les 4 variables, ou parmi d'autres variables que vous trouverez pertinentes, trouvez une variable dont la loi est normale.
- **un test de comparaison de deux populations (dans le cas gaussien)** : choisissez 2 clusters parmi ceux que vous aurez déterminé. Sur ces 2 clusters, testez la variable gaussienne grâce à un test de comparaison.

Livrables attendus

- Le dendrogramme dans un format image (.png ou .jpg)
- Le code Python ou R reprenant tous les éléments de l'énoncé.
- Un fichier CSV contenant la liste des pays ainsi que (pour chacun d'entre eux) leur groupe que vous aurez déterminé après avoir découpé le dendrogramme.
- Un fichier CSV contenant les centroïdes des groupes et leurs coordonnées dans chacune des dimensions.
- Une présentation sous forme de slides pour la soutenance.

Modalités de la soutenance

Vous présentez vos résultats à votre hiérarchie, qui vous a demandé une présentation sur des "slides".

Rappelez en introduction les enjeux, expliquez dans le détail votre démarche (sans avoir peur de rentrer dans les détails mathématiques), et présentez vos résultats et recommandations.

A vous de bien équilibrer les parties de votre présentation, qui doit durer 20 minutes suivies de 10 minutes d'éventuelles questions de votre hiérarchie.

Ressources complémentaires

<https://www.youtube.com/watch?v=uV5hmpzmWsU>

- Vidéo très didactique d'introduction à l'ACP.

<http://www.jybaudot.fr/Analdonnees/factorielles.html>

- Cours qui explique l'analyse factorielle (quasiment) sans formules.

http://factominer.free.fr/index_fr.html

- Factominer est un package phare en R spécialisé dans l'analyse exploratoire multivariée (ACP, CAH, etc.) Retrouvez un exemple d'ACP avec Factominer à [cette adresse](#).

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/117-hcpc-hierarchical-clustering-on-principal-components-essentials/>

- Tutoriel en anglais de clustering avec Factominer.

<http://www.grappa.univ-lille3.fr/~ppreux/ensg/miashs/fouilleDeDonneesII/tp/segmentation-hierarchique/>



<https://datascienceplus.com/k-means-clustering-in-r/>

- Si vous travaillez avec R, regardez le code de ces deux ressources.

Compétences évaluées



Interpréter une ACP

-  Tester l'adéquation à une loi par un test statistique
-  Construire et lire un dendogramme

OpenClassrooms

L'entreprise
Alternance
Forum
Blog
Nous rejoindre

Entreprises

Employeurs

En plus

Devenez mentor
Aide et FAQ
Conditions Générales
d'Utilisation
Politique de Protection
des Données
Personnelles
Nous contacter

 Français 

