

Étude de données sur la santé publique



**Food and Agriculture
Organization of the
United Nations**

Contexte de la mission.

Ce projet nous emmène en tant que Data Analyste au sein d'une équipe de chercheurs de la Food and Agriculture Organization of the United Nations (FAO).

C'est l'un des organes qui composent l'ONU et dont l'objectif est "d'aider à construire un monde libéré de la faim". Le responsable de l'équipe, Marc, un chercheur en économie de la santé, est enthousiaste à l'idée de votre arrivée. En effet, Julien, l'ancien Data Analyst, a récemment été muté, et votre équipe vient de se voir confier la réalisation d'une étude de grande ampleur sur le thème de l'alimentation et plus particulièrement sur la sous-nutrition dans le monde.

A ce titre, Marc nous fait part du contexte et des enjeux du projet, il s'agit d'une grosse étude sur l'alimentation dans le monde. Notre rôle dans ce projet est de compléter les analyses de l'ancien Data analyste qui a traité les données de 2018 à maintenant. Nous devons donc nous occuper de la partie « historique » de 2013 à 2017.

Ressources Disponibles.

Nous avons à disposition 4 fichier csv :

1. Un fichier de disponibilité alimentaire pour l'année 2017 qui contient

17 variables :

- Le pays.
- Le produit.
- 4 variables de disponibilité alimentaire.
- 10 variables qui précisent les utilisations/moyens d'acquisition de la disponibilité alimentaires, exprimées en milliers de tonnes.

- Enfin, une variable type qui précise si le produit est d'origine animale ou végétale.

2. Un fichier d'insécurité alimentaire, qui inclut :

- Le pays.
- Intervalle d'années.
- Le nombre de personnes en sous-alimentation en millions d'habitants (ex. : pour l'intervalle 2012-2014, le nombre de personnes correspond à la moyenne sur ces 3 années, on peut le résumer en 2013).

3. Un fichier de population qui inclut :

- Le pays.
- L'année.
- La population du pays, exprimée en milliers d'habitants.

4. Un fichier d'aide alimentaire qui contient :

- Le pays bénéficiaire.
- L'année.
- Le produit.
- La quantité qui a été donnée comme aide alimentaire, en tonnes.

Renommage et formatage des Données préalable aux analyses.

On commence ce projet par une inspection des données, leur type ainsi que leurs formats.

On remarque que certaines données contiennent des accents, espace inutile, et que certaines valeurs ne sont pas exprimées dans la même échelle d'un fichier à l'autre.

Il est donc impératif d'harmoniser nos structures de données ainsi que les données qui y sont contenues afin d'assurer la cohérence et la fiabilité de nos analyses futures.

On commence donc par le renommage des valeurs qui ne correspondent pas à nos standards d'harmonisation.

Dans Population : Valeur => population

Dans sous nutrition : Valeur => sous_nutrition

On convertit les données stocker dans ses colonnes en type numérique.

Dans disponibilité : toutes les colonnes afin d'y supprimer les accents, majuscules et de simplifier certains termes (« Disponibilité alimentaire en quantité (kg/personne/an → disponibilite_kg_p_an » par exemple).

Remplacement sur les 4 fichiers de tous les accents, lettre en majuscule par leur équivalent minuscule, et on supprime les espaces inutiles.

Conversion des colonnes devant recevoir des quantités en milliers de tonnes en kilos afin d'harmoniser nos différents fichiers (avec un filtre < 1 000 000, afin de détecter les saisies déjà faites en kilos).

L'annexe fournis nous indique également que les périodes d'année pour nos données correspondent à l'année « d'entre deux » (ex : 2012-2014 = 2013), on doit donc indiquer à notre dataframe l'équivalence des périodes avec les bonnes années qui y correspondent.

Jointure des fichiers pour l'analyses des données.

On crée un dataframe via une jointure entre les tables sous-nutrition et population sur les champs 'zone' et 'annee' en outer pour garder les lignes où un champ est manquant.

On va ensuite faire un filtre sur l'année 2017, mais avant on s'assure que la colonne année est bien en numérique avant de faire ce filtre sur 2017.

On peut ensuite calculer le nombre de personnes en sous nutrition sur cette année, et la proportion par rapport à la population mondiale en 2017.

On joint les dataframe dispo_alimentaire et le dataframe sur l'année 2017 pour y rajouter la population.

Calcul grâce à la dispo_kcal_p_j et à la population la quantité de ressources en kcal disponible mondialement en 2017 mais aussi le nombre de personnes qu'on peut nourrir selon la provenance de la nourriture, calculer la disponibilité totale et la proportion de l'utilisation de cette ressources (pertes, alimentation du bétail, alimentation de la population)

Création d'une liste de produit correspondant aux céréales afin de suivre l'utilisation des différents types de céréales.

Suivis du taux de sous nutrition dans le monde selon le pays, puis suivis de l'évolution de l'aide alimentaire par pays sur les années de 2013 à 2017.

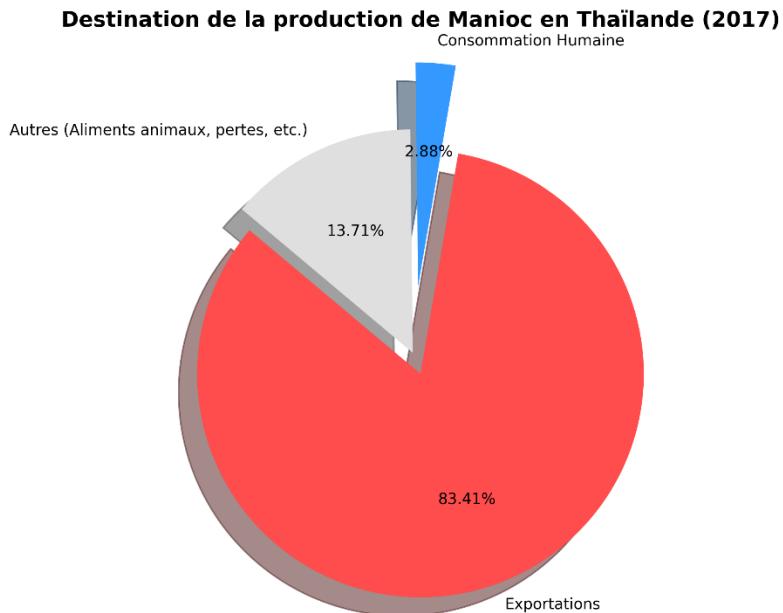
Définition d'un palmarès (top et flop) des disponibilités par habitant selon le pays.

On va finir par étudier l'étude de cas de la Thaïlande, qui est un parfait exemple des problèmes qu'on a pu détecter, gros exportateur de manioc, produit d'alimentation de bases, mais qui a une population qui souffre d'un problème de sous-alimentation.

Marc nous propose de faire d'autre analyses complémentaires, j'ai décidé de créer un dictionnaire géographique afin de pouvoir étudier les problèmes de sécurité alimentaire selon leur zones géographique et de mettre en avant les problèmes majeurs rencontré dans ses pays concernés.

J'ai aussi décidé de calculer la ration des pertes mondiales afin d'alerté que l'aide alimentaire disponible est ridiculement faible en comparaison à ce qu'on perd.

Concernant l'étude de cas sur la Thaïlande j'ai réalisé un graphique sectoriel afin de visualiser la répartition de la production de manioc avec matplotlib



Pour finir j'ai d'abord trié les céréales pour récupérer les 3 céréales les plus produites dans le monde (Blé, Maïs, Riz) pour ensuite calculer la proportion de leur utilisation dans l'alimentation (soit animale, soit humaine)

Conclusion du projet.

Les différentes analyses m'ont permis d'identifier les différents facteurs de la faim dans le monde. J'ai montré grâce à ces analyses que le problème de faim dans le monde n'est pas un problème de production mais un problème de distributions et de logistique. Nous avons montré que les pertes mondiales représentent un volume tellement colossal qu'une simple réduction de 1.5% permettrait de subvenir au besoin alimentaire mondial.