

Projet : Analyse de la productivité agricole en Afrique de l'ouest

Par

Gérard De La Paix BAYIHA

Table des matières

| | | |
|---------|---|---|
| 1 | Contexte :..... | 3 |
| 2 | Objectif général : | 3 |
| 2.1 | Objectifs spécifiques | 3 |
| 2.1.1 | Utiliser Python pour faire l'Analyse exploratoire de la base de données..... | 3 |
| 2.1.2 | Utiliser Mysql pour faire certaines opérations..... | 3 |
| 3 | Etape 1 : Analyse exploratoire de la base de données..... | 3 |
| 3.1 | Les noms des variables..... | 4 |
| 3.2 | La nature et le nombre de types de variables | 4 |
| 3.3 | Origine des données | 4 |
| 3.4 | Représentativité des indicateurs pour chacun des pays | 5 |
| 3.5 | Représentativité des cultures pour chacun des pays | 5 |
| 4 | Etape 2 : Opérations à travers MySQL..... | 6 |
| 4.1 | Calculer le rendement moyen par culture pour la région de l'Afrique de l'ouest..... | 7 |
| 4.1.1 | Calculer le rendement moyen par culture et pour chacun des pays de la sous-région | 7 |
| 4.1.1.1 | Maïs..... | 8 |
| 4.1.1.2 | Tomate, fraîches | 8 |

1 Contexte :

Nous disposons d'une base de données de l'Afrique de l'ouest qui couvre la période 2000 à 2022. Elle regroupe les éléments suivants :

- Domaine : Culture
- Indicateurs : superficie récoltées(ha), production (tonnes), rendement (t/ha)
- Zone : Afrique de l'ouest (les pays)
- Cultures : maïs, riz, tomate, pomme de terre, sorgho
- Source : FAOSTAT,2025

2 Objectif général :

Combiner l'utilisation de Python et un système de gestion de bases de données (MySQL)

2.1 Objectifs spécifiques

Nous avons deux objectifs spécifiques.

2.1.1 Utiliser Python pour faire l'Analyse exploratoire de la base de données

Il s'agira de traiter les données après cette analyse, dans le cadre d'un autre projet. L'objectif ici est de présenter un rapport à travers le module `AutoViz_Class()`. Dans ce document, nous mettrons uniquement en évidence les résultats de l'analyse qui influenceront les opérations décrites dans le **sous-objectif 2.1.2.**

2.1.2 Utiliser Mysql pour faire certaines opérations

A travers MySQL, nous souhaitons avoir sur la période 2000 - 2022 :

- Calculer le rendement moyen par culture et par pays
- Repérer les pays ayant augmenté leur production entre deux années
- Comparer la productivité entre culture de même famille : ici les céréales
- Créer des vues pour suivre les top 5 des cultures par pays

3 Etape 1 : Analyse exploratoire de la base de données

La base de données, issue de FAOSTAT, couvre la période 2000-2022. Elle contient des informations sur les superficies, rendements et productions pour le maïs, le riz, la tomate, la pomme de terre et le sorgho dans 16 pays d'Afrique de l'Ouest : Bénin, Burkina Faso, Cabo Verde, Côte d'Ivoire, Gambie, Ghana, Guinée, Guinée-Bissau, Libéria, Mali, Mauritanie, Niger, Nigéria, Sénégal, Sierra Leone, Togo.

Nous avons reformulé les titres des colonnes et ajouté 4 colonnes relatives aux identifiants (pays, année, produit, indicateur/élément). Quatre colonnes (code domaine, code élément, code année, code produit) jugées inutiles pour notre analyse ont été supprimées.

Après l'analyse exploratoire, nous obtenons les résultats ci-dessous.

Après avoir lancé l'analyse exploratoire, nous avons les résultats suivants :

3.1 Les noms des variables

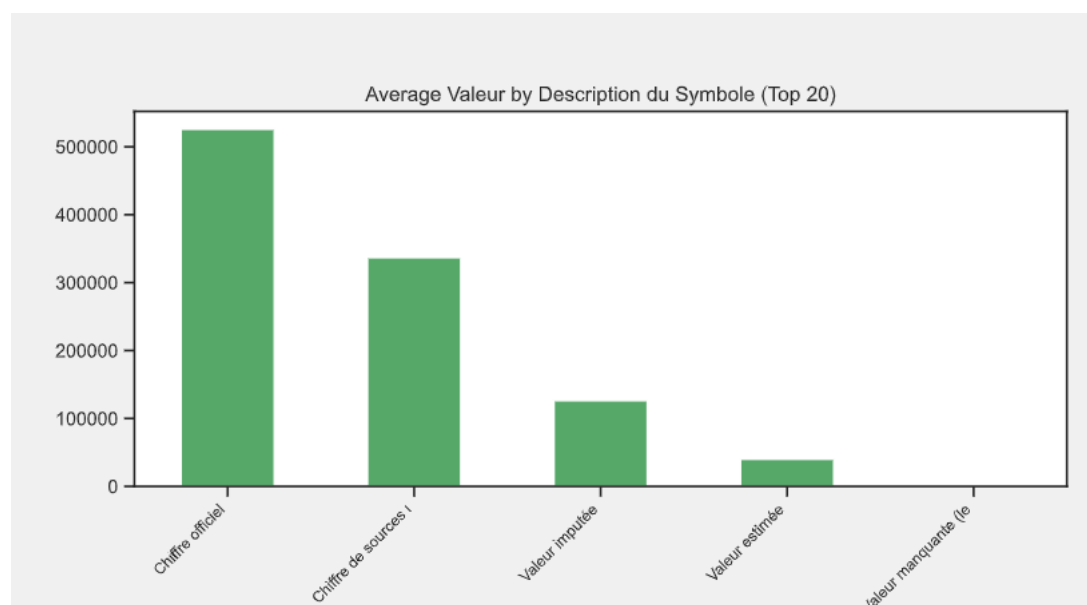
Nous avons 4820 observations et 14 variables. Les variables sont :

- **Code Domaine** : donne une forme d'identifiant au domaine
- **Domaine** : elle représente l'ensemble des domaines. Elle renvoie aux productions végétales et animales
- **Nom_Pays** : elle donne la liste des pays pris en compte
- **Nom_Indicateurs** : elle renvoie à l'ensemble des indicateurs pris en compte dans la base. Il y en a trois : rendement, production, superficie cultivée
- **Année** : renvoie à l'ensemble des années considérées dans l'étude
- **Nom_Produit** : elle liste toutes les cultures de la base par pays
- **Unité_Indicateur** : elle porte sur les unités de chaque indicateur : rendement(t/ha), production (t) et superficie (ha)
- **Valeur** : elle présente les valeurs numériques de chaque indicateur
- **Symbole** :
- **Description** : elle indique l'origine des informations (source officielle de l'Etat, source des organisations internationales)
- **ID (indicateur, produit, année, valeur)** : Ils ont été auto-incrémentés

3.2 La nature et le nombre de types de variables

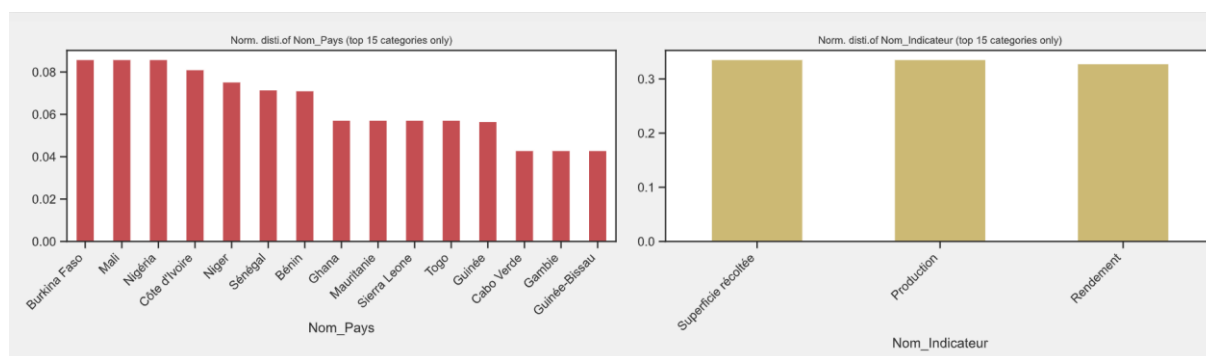
| Nature de la variable | Nombre | Nom de la/des variable(s) |
|-----------------------|--------|------------------------------------|
| Integer | 1 | Année |
| Objet | 8 | Reste des variables |
| Float | 1 | Valeur |
| Integer (ID) | 4 | Indicateur, produit, année, valeur |

3.3 Origine des données



La figure ci-dessus nous montre que les données proviennent de cinq sources : officielles, d'autres sources, des imputations (c'est-à-dire que l'organisation a remplacé certaines valeurs manquantes par un nombre spécifique), des valeurs manquantes qui ont été estimées et enfin quelques valeurs manquantes non traitées. Sur une échelle de 0 à 500 000, la moyenne des données issues de sources officielles est de 500 000, suivi d'autres sources de données (études, rapports de projets). En conclusion, nous pouvons dire que ces données sont fiables.

3.4 Représentativité des indicateurs pour chacun des pays



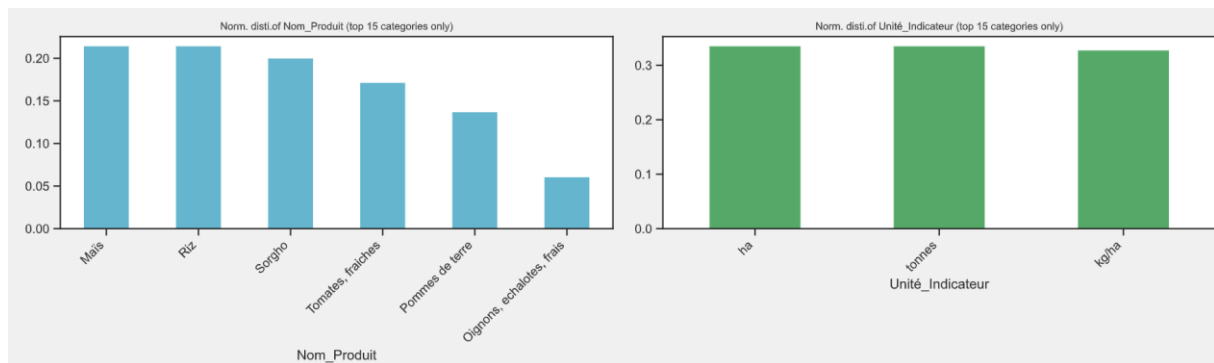
La figure de gauche présente, en abscisse, les pays de l'Afrique de l'ouest inclus dans l'analyse et en ordonnée, le poids que chacun occupe dans la base de données. La somme des poids de chacun des pays doit donner 1.

Il apparaît que la base de données contient proportionnellement plus d'informations (jusqu'à 0.8) pour le Burkina-Faso, le Mali, le Nigéria que pour d'autres tels que la Gambie, de la Guinée-Bissau. Cette différence de représentativité implique l'analyse des rendements moyens à l'échelle de la sous-région pourrait contenir un biais. Il n'est pas conseillé de calculer un rendement moyen agrégé pour ces cultures dans la région d'Afrique de l'Ouest. Il est donc préférable de :

- Privilégier des calculs détaillés par culture et par pays.

La figure de droite illustre, via un diagramme en barres, que les trois indicateurs étudiés présentent des poids de représentativité similaires dans l'ensemble des données. Par conséquent, aucun indicateur ne devrait influencer disproportionnellement les résultats.

3.5 Représentativité des cultures pour chacun des pays



La figure de gauche montre que le maïs, le riz et, dans une moindre mesure, le sorgho sont plus représentés dans la base de données que des cultures comme la pomme de terre ou l'oignon.

Cette répartition facilite l'analyse, bien que ces cultures appartiennent à des familles différentes. Ainsi :

- Une analyse agrégée des céréales comme le maïs et le riz est envisageable à l'échelle régionale ;
- En revanche, le sorgho, bien qu'étant aussi une céréale, ne devrait pas être intégré à cette analyse globale en raison de sa faible représentativité.

Comme précédemment, une analyse individuelle par produit reste la plus pertinente.

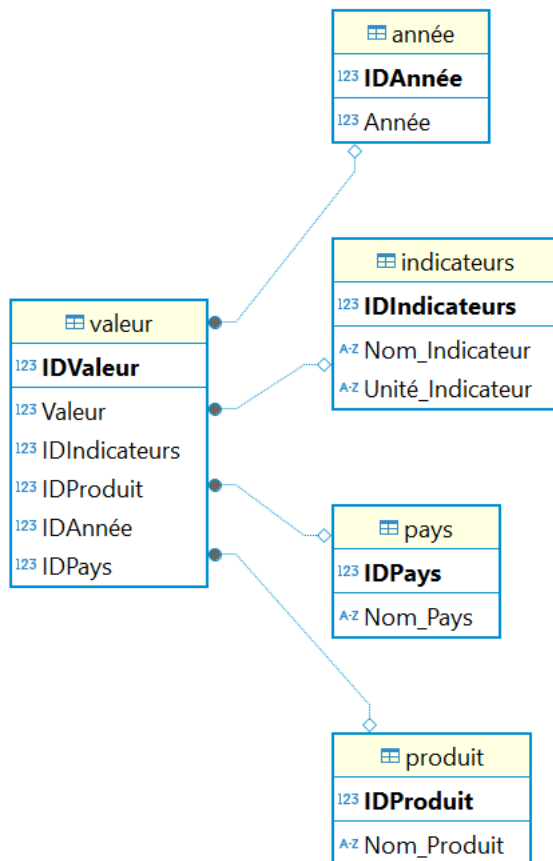
4 Etape 2 : Opérations à travers MySQL

En examinant les colonnes *Indicateurs*, *Unité* et *Valeur*, on constate que les trois indicateurs : rendement (t/ha), superficie récoltée (ha) et production (tonne) sont regroupés dans une même colonne intitulée : valeurs.

Pour atteindre les objectifs d'analyse, il est nécessaire de modéliser la base afin d'éviter cette structure combinée.

L'approche retenue est la suivante :

- Création d'un modèle conceptuel de données (MCD) définissant plusieurs tables, chacune correspondant à une entité avec ses attributs spécifiques.
- Mise en place d'un modèle logique de données (MLD) pour relier ces tables et faciliter les requêtes SQL (figure ci-dessous).



L'objectif est d'obtenir une structure où chaque observation possède un seul indicateur, une unité et une valeur spécifique, plutôt qu'un mélange.

Réponses aux questions

4.1 Calculer le rendement moyen agrégé des cultures pour la région de l'Afrique de l'ouest

Le calcul de ce rendement comporte un biais lié à la représentativité inégale des pays comme indiqué dans le point 3.4. Pour régler ce problème, il faut passer par une analyse individuelle par culture et par pays. A titre d'exemple, Nous prendrons ainsi l'exemple du Maïs et de la tomate pour chacun des pays.

4.1.1 Calculer le rendement moyen par culture et pour chacun des pays de la sous-région

Nous allons donner pour deux cultures (Maïs, tomate), la liste des trois pays sur les 16 en Afrique de l'Ouest ayant les rendements moyens les plus élevés sur la période 2000-2022 et conclure. Il est important de rappeler que pour ces deux cultures, sur les 16 pays d'Afrique de l'Ouest, 15 en produisent de manière conséquente. Nous avons effectué des requêtes sur MySQL (voir le script).

4.1.1.1 Maïs

Tableau 1 : liste des trois pays ayant les rendements moyens les plus élevés, associés à leurs écart-types (ET) et la position occupée en Afrique de l'Ouest par rapport aux ET sur la période 2000-2022

| Pays | Unité_Indicateur | Rendement_moyen | Ecart-type (ET) | Rang (Fonction de l'ET) |
|---------------|------------------|-----------------|-----------------|-------------------------|
| Mali | kg/ha | 2179.86 | 36,5 | 10 ^{ème} |
| Cote d'Ivoire | kg/ha | 2081.82 | 45.03 | 15 ^{ème} |
| Ghana | kg/ha | 1854.21 | 38,17 | 12 ^{ème} |

Source : FaoStat, 2025

Les résultats montrent que le Mali, la Côte d'Ivoire et le Ghana affichent les rendements moyens les plus élevés. Cependant, il est important de savoir si les rendements pour ces trois pays sont stables (homogènes) au cours de cette période ou alors, si le fait d'avoir des rendements moyens élevés, seraient dus à des valeurs extrêmes.

Pour ce fait, l'examen des écarts-types révèle que ces trois pays figurent aussi parmi ceux ayant la plus forte variabilité, signe que les bons rendements peuvent être liés à des années exceptionnelles plutôt qu'à une performance stable.

En comparaison, le Niger présente un rendement moyen plus faible (962,3 kg/ha) mais un ET de seulement 24,8, soit le plus faible de la sous-région, indiquant une meilleure stabilité. Pour les décideurs, cela signifie que le Niger offre un potentiel plus fiable pour améliorer la productivité via des innovations technologiques, tandis qu'en Côte d'Ivoire, la priorité devrait être de réduire la variabilité dont les causes peuvent être diverses.

4.1.1.2 Tomate, fraîches

Le même raisonnement est fait pour la tomate qui est une culture de base pour la sécurité alimentaire des pays d'Afrique de l'Ouest.

Tableau 2 : liste des trois pays ayant les rendements moyens les plus élevés, associés à leurs écart-types (ET) et la position occupée en Afrique de l'Ouest par rapport aux ET sur la période 2000-2022

| Pays | Unité_Indicateur | Rendement_moyen | Ecart-Type | Rang (Fonction de l'ET) |
|---------|------------------|-----------------|------------|-------------------------|
| Sénégal | kg/ha | 22685.91 | 33.33 | 6 ^{ème} |
| Niger | kg/ha | 20878.87 | 24.8 | 1 ^{ère} |
| Mali | kg/ha | 17659.17 | 36.5 | 10 ^{ème} |

Source : FAOSTAT, 2025

Pour la tomate, le Sénégal est en tête, suivi du Niger et du Mali.

Comme pour le maïs, le Niger se distingue par la stabilité de ses rendements (et le plus bas de la région), ce qui en fait une cible stratégique pour les investissements (par exemple, innovation technologique) visant à accroître la productivité.

Les résultats dans **le tableau 2** montrent que les trois pays ayant les rendements moyens les plus élevés sont le Sénégal avec 22 685,91 kg/ha, suivi du Niger avec 20 878,87 kg et le Mali avec 17 659 kg. Comme précédemment, nous calculons les écarts types des pays de la sous-région pour mieux situer les rendements de ces pays en termes de stabilité sur la période 2000 à 2022.

1. Identifier les pays ayant augmenté leur production entre deux années

A venir