



Anomalies des interventions du 01-01-2023 au 15-03-2024

Soufiane BOUTASKNITE
Stage de fin d'étude
Tutrice : Maryam Guessous

I. Les Valeurs Manquantes :

Pour chaque colonne, nous calculons le pourcentage de valeurs manquantes. Le tableau suivant ne présente que les colonnes ayant un pourcentage de valeurs manquantes différent de zéro.

Colonne	Pourcentage des cases vides
Culture	94.83 %
Totale (L)	50.52 %
Litres / h	52.93 %
Litres / ha	57.01 %
Heures Moteurs	50.52 %
CO2 Émis (Kg)	50.52 %
Temperature	50.52 %

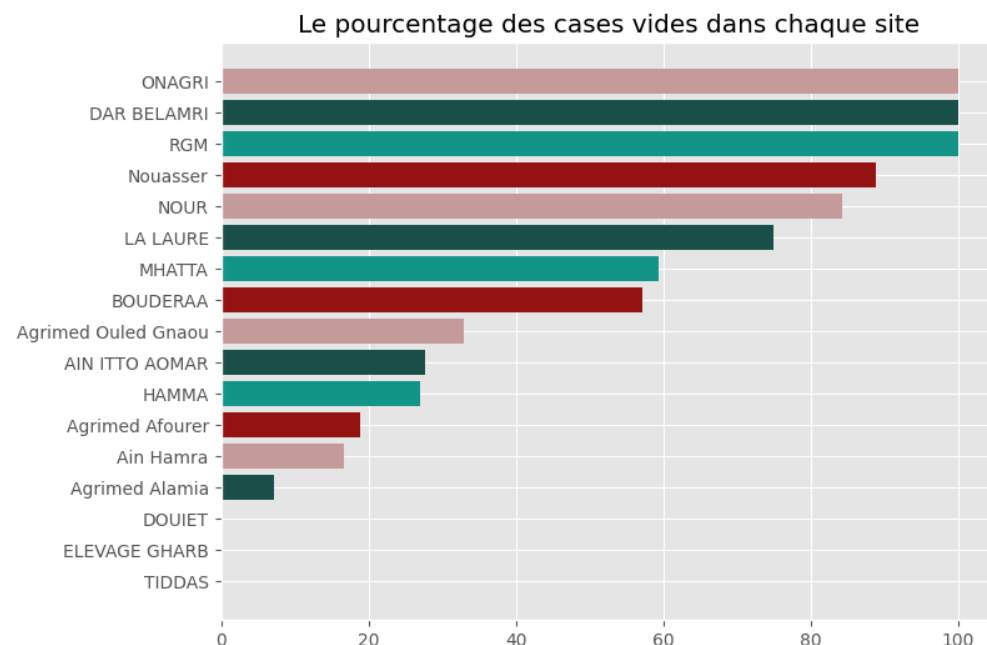
La colonne "Culture" a le pourcentage le plus élevé de cases vides avec 94,83 %, ce qui suggère que la majorité des données pour cette catégorie sont manquantes ou non enregistrées. Les autres colonnes, telles que "Totale (L)", "Heures Moteurs", "CO2 Émis (Kg)" et "Temperature", ont toutes un pourcentage identique de cases vides, soit 50,52 %.

L'uniformité du pourcentage de cases vides dans ces colonnes pourrait indiquer un problème systématique dans la collecte des données ou un critère de saisie qui n'a pas été respecté. La gestion des données manquantes est cruciale car elle peut affecter l'analyse et les conclusions tirées de ces données. Il est important d'enquêter sur la raison de ces valeurs manquantes et de déterminer si elles peuvent être récupérées ou si des méthodes d'imputation doivent être utilisées pour pallier ces lacunes.

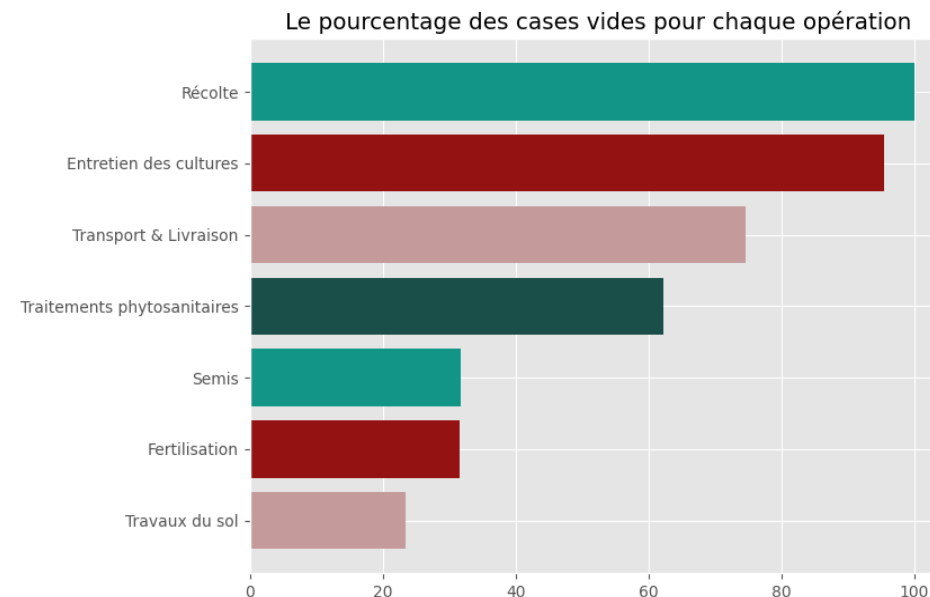
La question soulevée est la suivante : pourquoi les pourcentages pour les colonnes "Litres / h" et "Litres / ha" diffèrent-ils des autres colonnes, alors qu'ils devraient normalement être calculés à partir de la colonne "Totale (L)" ? Pour répondre à cette question, nous examinons les lignes où la colonne "Totale (L)" est non vide et la colonne "Litres / ha" est vide. Nous constatons qu'il y a 280 de telles lignes où "Totale (L)" est égal à zéro, ce qui implique des interventions sans consommation, une situation impossible. Cela suggère que ces lignes étaient initialement vides, mais ont été remplies par des zéros d'une manière ou d'une autre.

Donc le vrai pourcentage des cases vide pour "Totale (L)", "Litres / h" et "Litres / ha" est 57,01 %. Ça concerne aussi "CO2 Émis (Kg)" et "Heures Moteurs". Ce n'est pas le cas pour "Temperature".

Source des cases vides :



Pour chaque site, nous calculons le pourcentage de données de consommation de carburant non enregistrées. Nous constatons qu'aucune donnée de consommation n'a été enregistrée dans les sites "ONAGRI" "DAR BELAMRI" et "RGM". En revanche, aucun des sites "DOUIET", "ELEVAGE GHARB" et "TIDDAS" n'a de données de consommation manquantes. Pour les autres sites, nous avons calculé différentes proportions, que nous avons représentées dans le graphe ci-dessus.



Le graphique ci-dessus indique que 100% des interventions de "Récolte" n'ont pas été associées à leurs consommations correspondantes, suivies par "Entretien des cultures". En revanche, seulement 23% des interventions effectuées n'ont pas leurs consommations correspondantes.

II. Valeurs Illogiques :

Superficie Cultivée :

Après une vérification approfondie, il a été constaté que la seule valeur négative, -0.44, correspond à la parcelle "FW 21". En consultant la liste des parcelles, il a été remarqué que la superficie cultivée de cette parcelle n'est pas définie, et de quelque manière que ce soit, elle a été remplacée par -0.44 dans les données.

Est-il possible que certaines parcelles aient une superficie cultivée égale à zéro ?

Pourcentage :

Nous avons identifié 70 interventions avec un pourcentage supérieur à 1, ce qui est considéré comme inexplicable. Après avoir mené une investigation pour déterminer si ces anomalies sont liées à un site spécifique ou à une opération particulière, nous n'avons pas obtenu de résultats concluants. Dans ce contexte, il est nécessaire de se poser la question suivante : comment cette colonne est-elle calculée ? Cela nous permettra d'obtenir une explication sur l'origine des pourcentages qui dépassent 1.

Total Trajet :

Comme indiqué dans le tableau, la colonne "Total Trajet" contient des valeurs négatives. Pour identifier ces cas, nous avons appliqué un filtre qui a permis de détecter trois interventions ayant des durées de trajet totales négatives, à savoir :

- 1) Id : 231644 Total Trajet : -0.2 minutes
- 2) Id : 229981 Total Trajet : -1.8 minutes
- 3) Id : 229292 Total Trajet : -1.7 minutes

Total Trajet :

Après vérification, on trouve que la consommation totale du carburant contient une seule valeur négative (-2651,53) et correspond à l'intervention dont Id est **264362** et on la supprime par la suite.

III. Non-correspondance avec la base de données :

Machine Utilisée :

En consultant la liste des machines, nous avons remarqué que les machines **DEUTZ AGROSUN**, **LANDINI 90** et **New Holland TS90** ne sont pas enregistrées dans la base de données. Il est possible qu'elles soient enregistrées sous d'autres noms.

Outil :

Certains outils utilisés lors des interventions de la période sélectionnée n'ont pas été trouvés dans la liste des outils. Cela peut être dû au fait qu'ils ne sont pas enregistrés dans la base de données ou qu'ils sont enregistrés sous des noms différents.
Le tableau suivant les illustre.

Outils non enregistrés
COVER-CROP
PULVERISATEUR FEEDER
SEMOIRE 28 LIGNES AVEC HERS SOLA 22055
Cultivateur 28 dents
EPANDEUR D'ENGRAIS AMAZONE
STUBBLE PLOW
BINEUSES

Largeur :

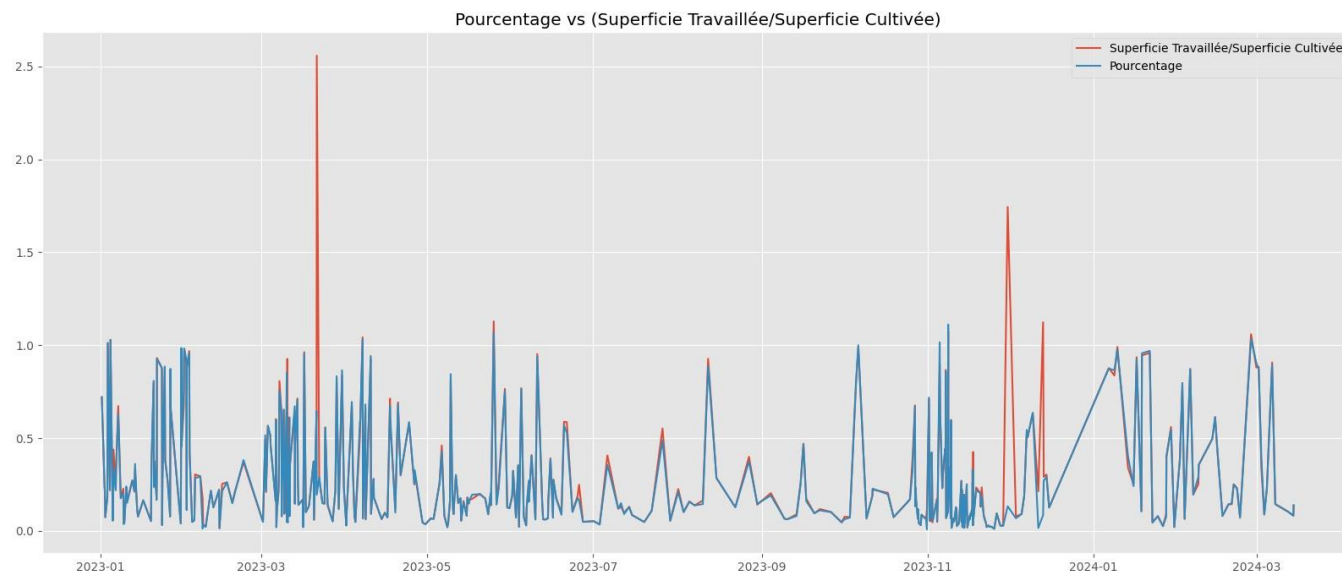
Après vérification, il a été constaté que les largeurs des outils **"Tondeuse Fairway 3 John Deere"**, **"Tondeuse Fairway 2 TORO"** et **"Cover crop 28 D"** ne correspondent pas à celles de la liste, ce qui soulève une question sur la cohérence des données relatives à cette caractéristique.

Site :

En consultant la liste des sites sur la plateforme **Traçagri**, il est observé que le site **"RGM"** est absent de la liste, ce qui demande une clarification.

VI. Relations non-établies :

✚ Pourcentage vs (Superficie Travaillée/Superficie Cultivée):

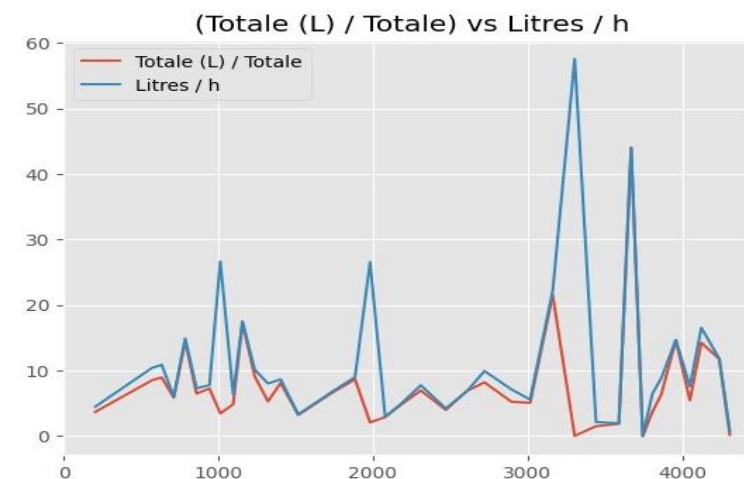


Nous constatons que dans certaines périodes, le rapport entre la superficie travaillée et la superficie cultivée n'est pas le même que celui calculé dans la colonne pourcentage.

De plus, le rapport lui-même atteint des valeurs supérieures à 1,

Ce qui soulève des questions sur la manière dont cette colonne est calculée. En répondant à cette question, nous pourrions identifier les causes de ces anomalies et comprendre plus en détail la logique sous-jacente à ces calculs.

✚ Comparaison de (Totale (L) / Totale) et Litres / h :



Le graphique ci-contre a été créé pour examiner si la consommation de carburant par heure est déterminée en divisant la consommation totale de carburant par la durée totale.

Les résultats indiquent que les deux courbes présentent une forme similaire pour la plupart des interventions. Cependant, les valeurs ne correspondent pas toujours, et il existe des périodes où il y a une grande différence entre les deux.

Cela nous amène à nous poser la question suivante : comment est calculée ou mesurée la consommation par heure ?

VI.Conclusion :

Nous avons identifié plusieurs anomalies qui doivent être corrigées avant de commencer notre analyse sur des bases solides. Voici un résumé de ces anomalies :

- 1) Les valeurs manquantes : Ce problème concerne principalement les valeurs de consommation. Pour le résoudre, nous avons identifié les sites avec le plus de valeurs manquantes, ainsi que les opérations associées.
- 2) Les valeurs illogiques : Les colonnes Superficie Cultivée, Totale Trajet et consommation totale du carburant présentent des valeurs négatives, tandis que la colonne Pourcentage comporte des valeurs supérieures à 1. De plus, nous avons observé des consommations nulles et des superficies nulles dans les données.
- 3) La non-correspondance avec la base de données : Les colonnes Outils, Site et Largeur contiennent des valeurs qui ne correspondent pas aux valeurs de la base de données.
- 4) Les relations mathématiques non-établie : C'est le cas de la colonne Pourcentage, qui ne correspond pas au rapport des deux superficies. De plus, la consommation par heure diffère de la consommation totale sur la durée.