

Lien GitHub : <https://github.com/NeoPgl/TP1-ATDN2>

RAPPORT TP1 - ATDN2

Etape 1 :

1)

Les variables disponibles sont :

- **surface_ha**
- **type_sol**
- **engrais_kg/ha**
- **precipitations_mm**
- **temperature_C**
- **rendement_t/ha**

Les variables correspondent aux colonnes du jeu de données csv.

2)

La ferme souhaite prédire le rendement du maïs en fonction de plusieurs facteurs. L'objectif est alors d'optimiser les ressources pour maximiser le rendement, ce qui permettra d'éviter des coûts inutiles et des pertes.

3)

La variable cible est 'rendement_t/ha', car c'est celle qu'on cherche à prédire et à optimiser. Les variables explicatives sont toutes les autres variables, correspondant aux colonnes du jeu de données (rendement t/ha pas pris en compte) (c.f. question 1, etape 1).

Etape 2 :

2.1)

Moyenne de rendement :

```
Moyenne du rendement : 7.378418687218944
```

Médiane du rendement :

```
Médiane du rendement : 7.349138167259971
```

La médiane est la valeur du milieu, soit **7.3**.

Mode du rendement :

Chaque valeur est unique, donc pas de mode. Selon certaines librairies pythons, La première valeur est considérée comme le mode dans le cas où chaque valeur est unique.

```
Mode du rendement : 3.000276469608442 (Fréquence : 1)
```

2.2)

Etendue :

```
Étendue du rendement : 8.995742859645505
```

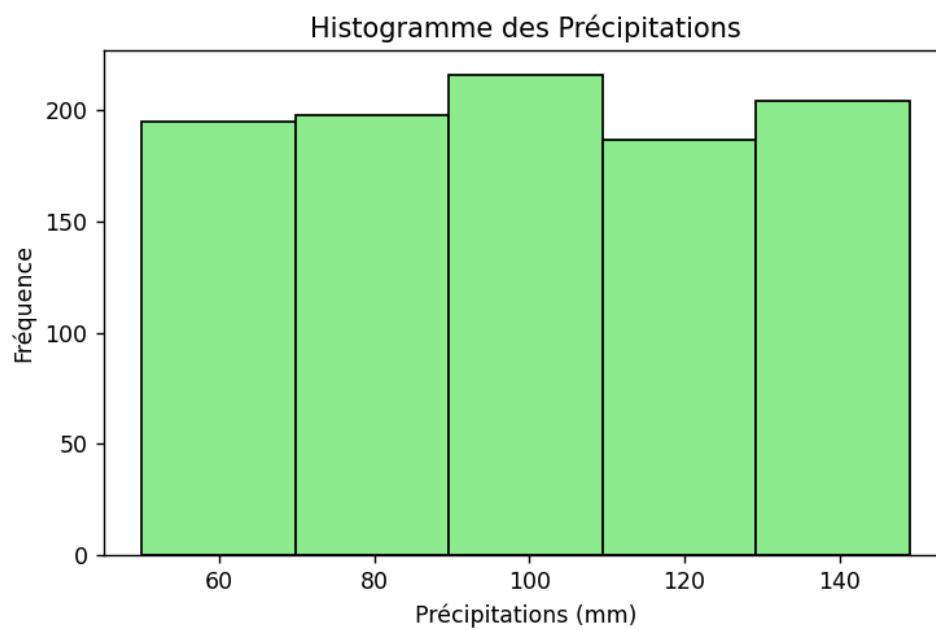
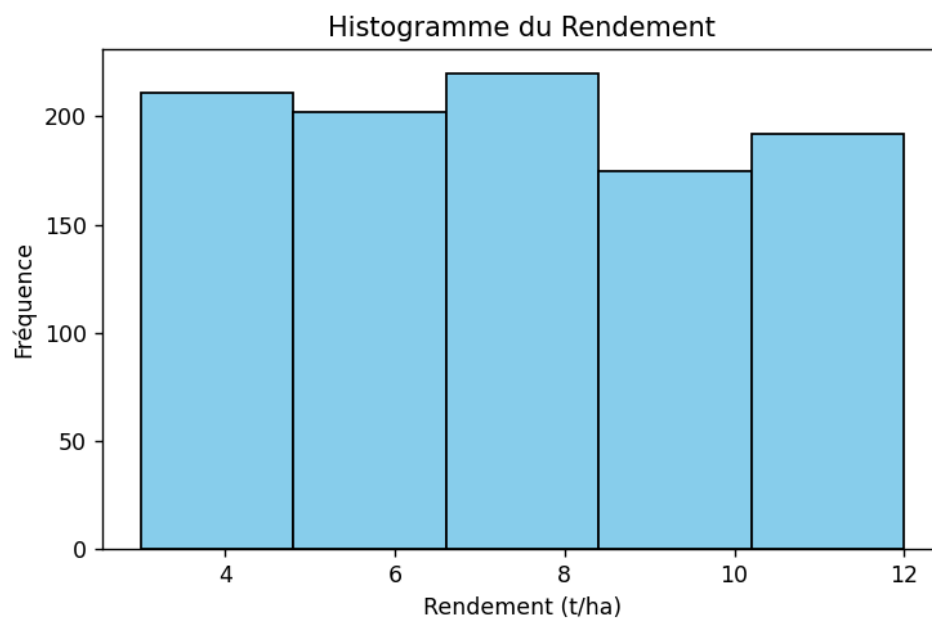
Variance :

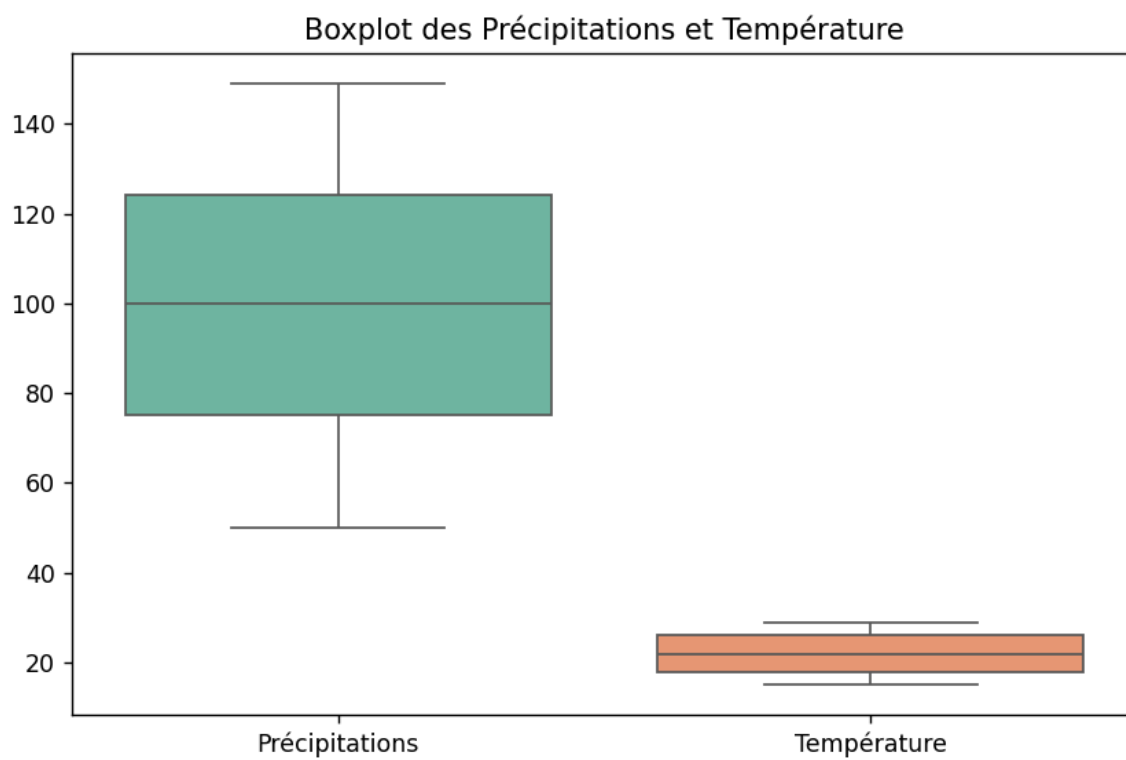
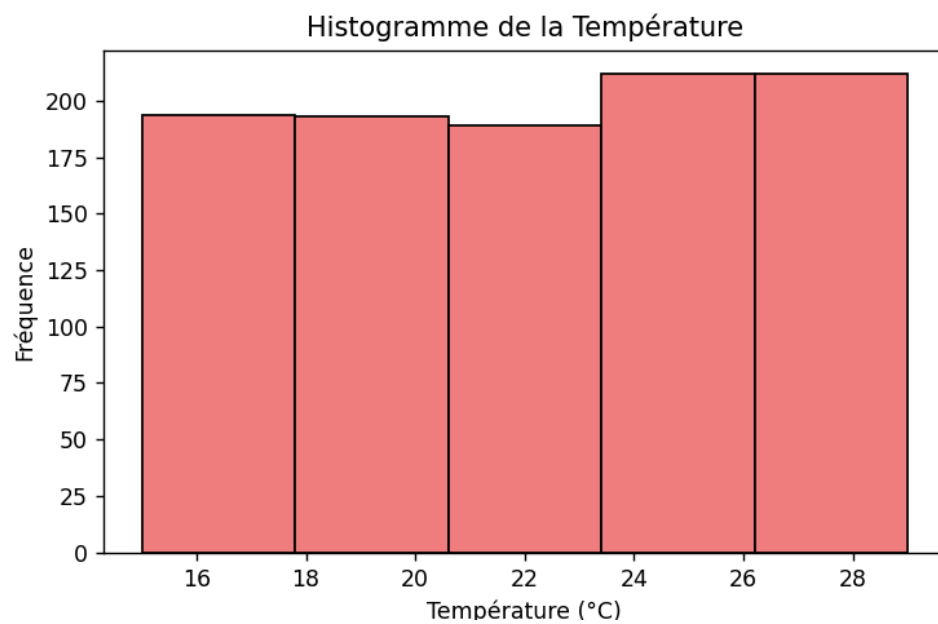
```
Variance du rendement : 6.5982488109958775
```

Ecart-type :

Écart-type du rendement : 2.5687056684244456

2.3)

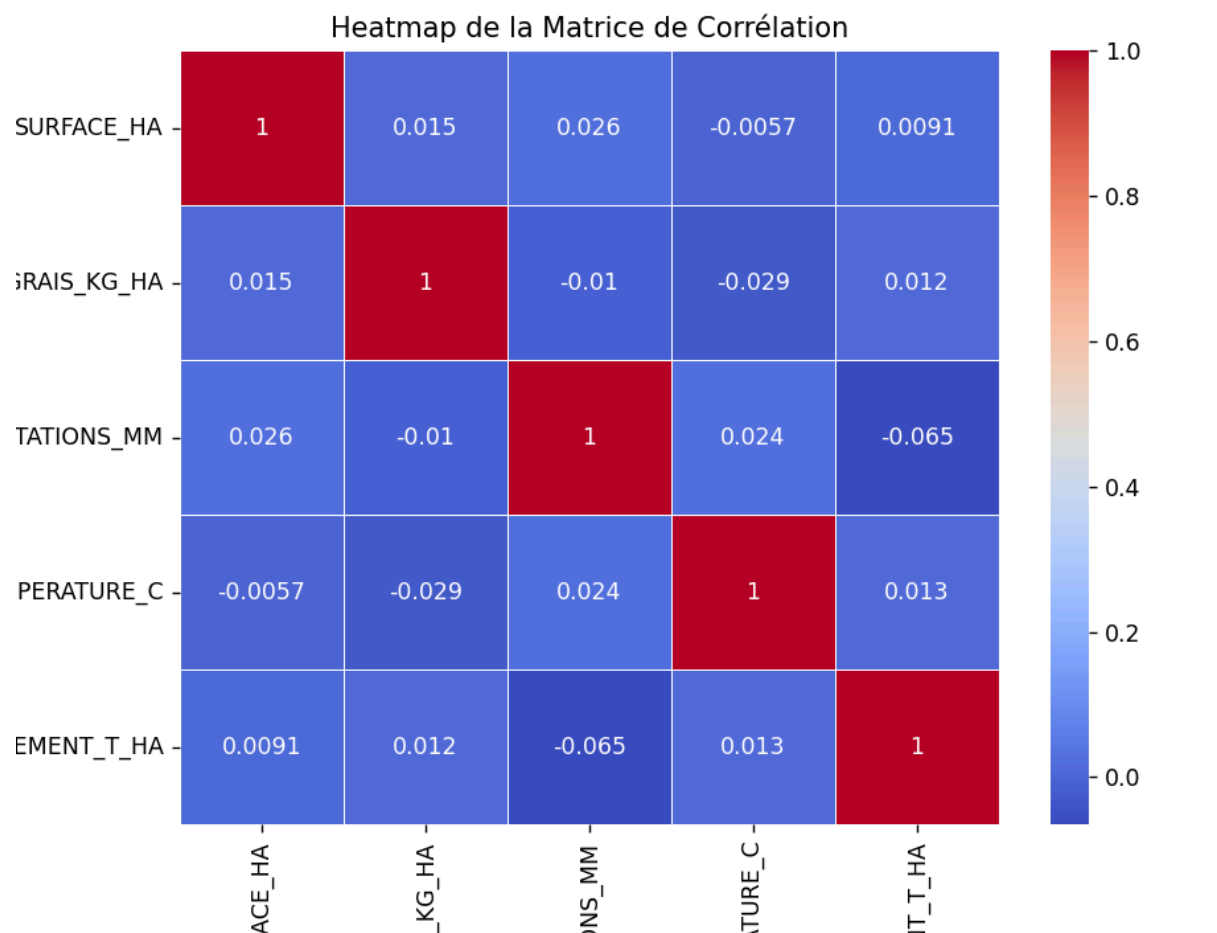




2.4)

Matrice de Corrélation :

	SURFACE_HA	ENGRAIS_KG_HA	PRECIPITATIONS_MM	TEMPERATURE_C	RENDEMENT_T_HA
SURFACE_HA	1.000000	0.014786	0.025900	-0.005717	0.009141
ENGRAIS_KG_HA	0.014786	1.000000	-0.010348	-0.028524	0.012368
PRECIPITATIONS_MM	0.025900	-0.010348	1.000000	0.023692	-0.065479
TEMPERATURE_C	-0.005717	-0.028524	0.023692	1.000000	0.013166
RENDEMENT_T_HA	0.009141	0.012368	-0.065479	0.013166	1.000000



La surface, l'engrais et la température sont les variables les plus élevées et qui semblent avoir le plus d'impact sur le rendement.

Etape 3 :

```
Test ANOVA détaillé :  
Moyenne totale du rendement : 7.378418687218944  
  
Moyennes par type de sol :  
TYPE_SOL  
Argileux      7.280166  
Limoneux      7.285699  
Sableux       7.565633  
Name: RENDEMENT_T_HA, dtype: float64
```

```
Somme des Carrés entre les groupes (SSB) : 17.90028686818752  
Somme des Carrés intra-groupe (SSW) : 6580.348524127689  
  
Degrés de liberté entre les groupes (df_between) : 2  
Degrés de liberté à l'intérieur des groupes (df_within) : 997  
  
Variance inter-groupe (MSB) : 8.95014343409376  
Variance intra-groupe (MSW) : 6.600148971040812  
  
Statistique F : 1.3560517305539492  
P-value : 0.2581509831874908
```

Le type de sol n'influence pas significativement le rendement ($p \geq 0.05$).

Etape 4 :

```
Évaluation des modèles :  
  
Régression Linéaire :  
MAE (Erreur absolue moyenne) : 2.06407924353965  
RMSE (Racine de l'erreur quadratique moyenne) : 2.426341503870645  
R2 : 0.0017138549149776638
```

```
Arbre de Décision :  
MAE (Erreur absolue moyenne) : 2.875673398767061  
RMSE (Racine de l'erreur quadratique moyenne) : 3.6015273386660347  
R² : -1.1995033159658535
```

La régression linéaire a l'air plus performante et adaptée car elle a un R^2 plus élevé que l'arbre de décision. De plus, elle a des valeurs pour les erreurs moins grandes.

Etape 5 :

Les variables sont importantes de par leur corrélation, les variables peuvent avoir un impact significatif sur les résultats et peuvent entraîner une augmentation ou une chute du rendement.

Adapter la ferme aux conditions climatiques, et choisir une bonne surface et un bon type de sol. Cela évitera des pertes et limitera les effets négatifs.

Choisir des modèles plus complexes pour avoir des résultats plus précis et cohérents. Le modèle linéaire peut se retrouver limité selon le contexte, un modèle non linéaire pourra mieux capturer les relations et les corrélations des données.

Pour optimiser sa production, la ferme pourra choisir un type de sol optimal et une surface optimisée.