

hogwarts

March 10, 2024

1 PROJET HOGWART

2 Hogwart (Poudlard) , Harry Potter et Hermione Granger, ça vous dis quelque chose ? si non, avant de commencer, je vous prie de cliquer sur ce lien : <https://fr.wikipedia.org/wiki/Poudlard>

Vous êtes plutôt Gryffindor (Gryffondor) ?

Hufflepuf (Poufsouffle) ?

Ravenclaw (Serdaigne) ?

Ou ... Slytherin (Serpentard) ?

Votre réponse ici —> : Serdaigne

Syntiche a reçu un appel urgent venant de Poudlard, c'est la panique , et c'est Ketsia qui prendra la relève !!!

En effet, Voldemort et son contingent n'ont point réapparu, et ainsi, cette préoccupation ne saurait troubler nos esprits. Toutefois, après un long et distingué service, le Choixpeau a pris la décision éminente de se retirer.

Le problème découle de la situation où le Choixpeau constituait l'unique méthode traditionnelle employée par l'école afin d'évaluer et d'améliorer le niveau de compétence de ses élèves parmi les diverses maisons : Gryffondor, Serpentard, Serdaigne et Poufsouffle.

Pour pallier à cette impasse, Dumbledore a conçu une solution aussi ingénieuse que téméraire : **le recours à un expert en analyse de donnée que tu es (oui, toi)**. Par une convergence des événements des plus fortuites, son choix s'est porté sur Ynov. Prodigieuse coïncidence ? (Je vous prie, jouez le jeu)

Dès lors, une vaste collecte de données a été orchestrée, impliquant 1600 élèves, soigneusement répartis dans les quatre maisons. Chacun d'entre eux s'est soumis à un examen de personnalité afin d'articuler son profil en fonction de dix caractéristiques primordiales.

- Courage (courage)
- Kindness (gentillesse)
- Lucidity (lucidité)
- Intelligence (intelligence)
- Creativity (créativité)

- Loyalty (loyauté)
- Cold blood (sang-froid)
- Sincerity (sincérité)
- Ambition (ambition)
- Wisdom (sagesse)

À chaque trait de personnalité (caractéristique ou colonne) correspond une évaluation sur une échelle de 10 points, exprimée en décimales, attribuée à chaque élève. Voici un exemple des résultats obtenus lors du test de personnalité pour un échantillon / étudiant :

Vous disposez de deux jeux de données `hogwart_train` (1) et `hogwart_test` (2) .

- Le premier jeux de données (1) : correspond aux élèves de l'année 2022-2023
- Le deuxième jeux de données (2) : correspond aux élèves de l'année 2023-2024

Votre objectif ? **faire preuve d'imagination** ...

Oui vous êtes Harry Potter ou Hermione Granger (à vous de choisir :)) et vous allez étudier les compétences des élèves, c'est-à-dire : analyser, observer, visualiser de fond en comble les données des deux années, et tirer des conclusions pertinentes.

je vous fais confiance

```
[2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

3 Partie 1 : Les qualités essentielles pour exceller en tant qu'élève (2022-2023)

Dans cette première partie, nous nous pencherons sur les caractéristiques clés qui définissent un bon élève à Poudlard.

En explorant le jeu de données (1) correspondant à l'année scolaire 2022-2023, nous chercherons à mettre en lumière une analyse approfondie des compétences et des points forts des élèves.

Pour se faire, tu vas répondre à cette liste de question :

0. ouvrir le dataset avec pandas et dire quelles sont les premières observations que tu fais sur ces données ?

```
[79]: Année_22_23 = pd.read_csv("hogwarts_22-23.csv", delimiter="\t")
Année_22_23.head()
```

```
[79]:      House  Courage  Kindness  Lucidity  Intelligence  Creativity  \
0  Slytherin  5.304088  3.185855  0.000000      2.678206      8.959938
1  Slytherin  4.398176  3.305061  4.721321      6.787622      7.244496
2  Slytherin  5.783194  4.314146  1.759267      4.059480      2.288574
```

3	Slytherin	4.151522	3.603175	4.149013	6.093137	2.529757
4	Slytherin	3.833501	3.496047	3.999172	5.166990	3.364945

	Loyalty	Cold blood	Sincerity	Ambition	Wisdom
0	5.849180	3.267036	3.736050	2.985840	3.599174
1	3.279486	5.257431	3.271450	3.483200	5.930750
2	5.102330	5.857933	5.859261	3.680916	3.776592
3	4.781199	4.550993	3.318560	3.654809	5.691006
4	6.652525	4.993203	3.996897	4.262987	1.954075

- les élèves sont ranger dans l'ordre des maison. On à aussi beaucoup de caractérisque différentes sur les élèves.

1. Comment les élèves sont-ils répartis (en nombre) dans les différentes maisons (Gryffondor, Serpentard, Serdaigle, Poufsouffle) ?

```
[80]: maison = Année_22_23['House'].value_counts()
maison
```

```
[80]: House
Slytherin      200
Gryffindor     200
Ravenclaw     200
Hufflepuff     200
Name: count, dtype: int64
```

- Les élèves sont répartis de manière équitables entre chaque maison
2. Comment les élèves sont-ils répartis (en nombre) dans les différentes maisons (Gryffondor, Serpentard, Serdaigle, Poufsouffle) en fonction de leurs qualités pour le courage, la gentillesse et l'intelligence ?

```
[81]: Année_22_23['COURAGE'] = Année_22_23['Courage'] > 5
Courage = Année_22_23.groupby('House')['COURAGE'].value_counts()
print(Courage, "\n")

Année_22_23['GENTILLESSE'] = Année_22_23['Kindness'] > 5
Kindness = Année_22_23.groupby('House')['GENTILLESSE'].value_counts()
print(Kindness, "\n")

Année_22_23['INTELLIGENCE'] = Année_22_23['Intelligence'] > 5
Intelligence = Année_22_23.groupby('House')['INTELLIGENCE'].value_counts()
print(Intelligence, "\n")
```

House	COURAGE	
Gryffindor	False	117
	True	83
Hufflepuff	True	125
	False	75

```

Ravenclaw  False      101
           True        99
Slytherin   False     103
           True        97
Name: count, dtype: int64

```

```

House      GENTILLESSE
Gryffindor  True        195
           False        5
Hufflepuff  True        179
           False        21
Ravenclaw   False     194
           True         6
Slytherin   False     199
           True         1
Name: count, dtype: int64

```

```

House      INTELLIGENCE
Gryffindor  True        133
           False        67
Hufflepuff  False     103
           True         97
Ravenclaw   False     128
           True         72
Slytherin   True        112
           False        88
Name: count, dtype: int64

```

- COURAGE: Les élèves sont répartis de manière plutôt équitable entre les différentes maisons. Cette qualité n'a donc aucune influence sur l'appartenance à la maison.
 - GENTILLESSE: On voit que les élèves de Gryffondor et de Poufsouffle sont presque tous au dessus de la moyenne en terme de gentillesse. Inversement, presque tous les élèves de Serdaigne et de Serpentard sont en dessous de la moyenne. Cette qualité est donc importante dans la répartition des élèves dans les maisons.
 - INTELLIGENCE: Les élèves sont répartis de manière plutôt équitable entre les différentes maisons. Les élèves de Gryffondor sont en moyenne plus intelligents et inversement pour les élèves de Serdaigne. Cette qualité ne doit pas être la plus importante dans la répartition des élèves dans les maisons.
3. Est-ce que les qualités (lucidité et intelligence) des élèves sont liées à leur appartenance (maison) ? Autrement, est-ce que le fait d'appartenir à une maison implique qu'on possède un peu plus ces 2 qualités ?

```

[82]: # On vérifie que la moyenne de ces deux qualités
      # soit supérieur à la moyenne (5).
      # Le "et" est matérialisé par le plus

```

```
Année_22_23['KindLucid'] = (Année_22_23['Intelligence'] +
↪Année_22_23['Lucidity']) / 2 > 5
kind_lucid = Année_22_23.groupby('House')['KindLucid'].value_counts()
kind_lucid
```

```
[82]: House      KindLucid
Gryffindor  True         147
           False         53
Hufflepuff  False        103
           True          97
Ravenclaw   False        117
           True          83
Slytherin   False        102
           True          98
Name: count, dtype: int64
```

- On voit que ces deux qualités sont importante pour entrer chez Gryffondor.
 - Les résultats sont plus omogènes pour les autres maison donc ces qualité non sont pas importante pour la répartition des élèves dans les autres maisons.
4. Quelles sont les qualités les plus communes chez les élèves de Gryffondor ?

```
[83]: Com_quality = Année_22_23.groupby('House').mean()
Com_quality.loc['Gryffindor'].sort_values() > 5
```

```
[83]: COURAGE      False
INTELLIGENCE      False
KindLucid         False
GENTILLESSE       False
Ambition          False
Courage           False
Creativity        False
Cold blood        False
Sincerity         False
Loyalty           True
Wisdom            True
Lucidity          True
Intelligence      True
Kindness          True
Name: Gryffindor, dtype: bool
```

- On voit que les qualité principales des élèves de Gryffindor sont: la gentillesse, l'intelligence, la lucidité, la sagesse et la loyauté
5. Moyennes des caractéristiques des élèves par maison :

Calculer la moyenne du courage, de l'intelligence, de la gentillesse, pour chaque maison (Gryffondor, Serpentard, Serdaigle, Poufsouffle) afin de déterminer les tendances centrales des caractéristiques

des élèves dans chaque maison.

```
[84]: Moyenne = Année_22_23.groupby('House')[['Courage', 'Kindness', 'Intelligence']].  
      ↪mean()  
      Moyenne
```

```
[84]:
```

	Courage	Kindness	Intelligence
House			
Gryffindor	4.472186	6.690932	5.870234
Hufflepuff	5.521260	6.008921	5.027867
Ravenclaw	4.979758	3.339592	4.367173
Slytherin	4.987841	3.432857	5.303907

3.0.1 Tendances centrales des maisons:

- Gryffondor: Les élèves sont très gentils et ils ont une intelligence supérieur à la moyenne mais ils ont un courage inférieur à la moyenne.
- Poufsouffle: Les élèves sont très gentils et ils ont un courage et une intelligence supérieur à la moyenne.
- Serdaigle: Les élèves ne sont pas très courageux, gentils et intelligent en moyenne.
- Serpentard: Les élèves ont un courage et une intelligence autour de la moyenne mais ils ne sont pas gentils.

6. Moyenne générale des caractéristiques des élèves :

Calculer la moyenne globale du courage, de l'intelligence, de la gentillesse, pour tous les élèves afin de comprendre les tendances centrales des caractéristiques des élèves à l'échelle de l'école dans son ensemble.

```
[85]: Moy = Année_22_23[['Courage', 'Kindness', 'Intelligence']].mean()  
      Moy
```

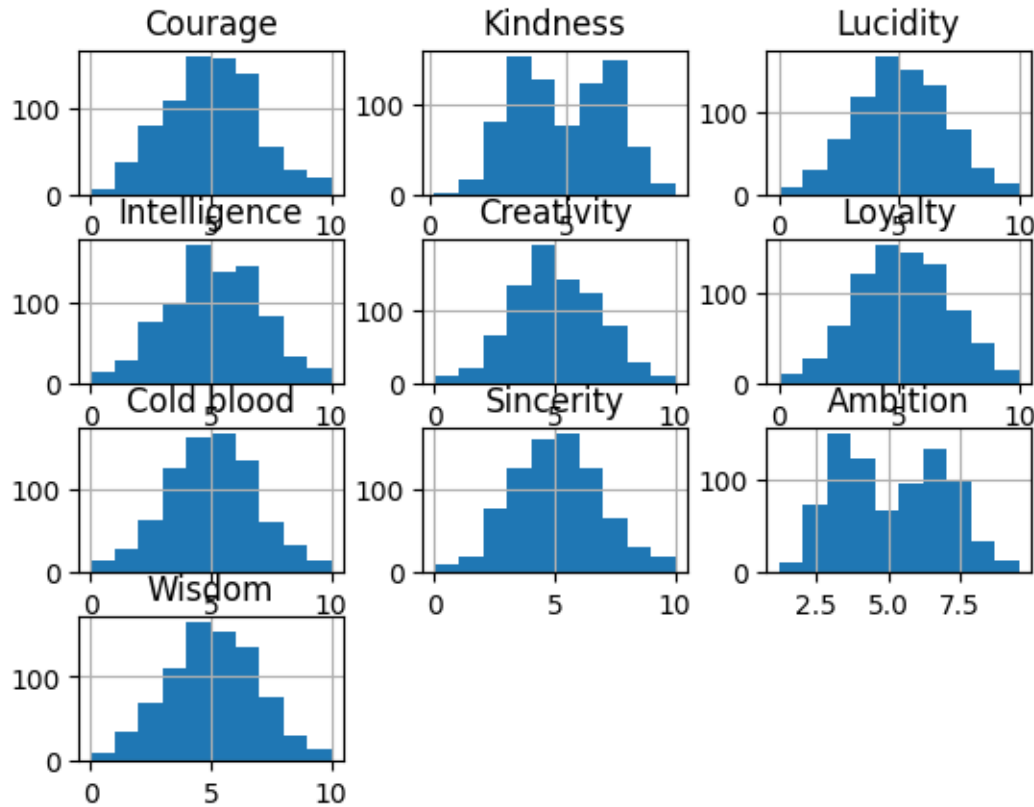
```
[85]: Courage          4.990261  
      Kindness         4.868076  
      Intelligence     5.142295  
      dtype: float64
```

- Il n'y a pas de tendance spéciale qui se dégage particulièrement. On voit que la moyenne de l'école pour ces trois caractéristiques est autour de la moyenne.

7. Analyse de la distribution des caractéristiques :

Utiliser des histogrammes pour visualiser la distribution des caractéristiques des élèves (par exemple, le courage) et identifier si elles suivent une distribution normale ou non.

```
[86]: histo = Année_22_23[['Courage', 'Kindness', 'Lucidity', 'Intelligence',  
      ↪'Creativity', 'Loyalty', 'Cold blood', 'Sincerity', 'Ambition', 'Wisdom']].  
      ↪hist()
```



- On voit que la plus part des caractéristiques sont réparties normalement: courage, lucidité, intelligence, créativité, loyal, sang-froid, sincérité, sagesse. Les graphiques épousent la courbe de Gause.
- Inversement, la gentillesse et l'ambition ne suivent pas cette courbe. Ces deux caractéristiques ne sont pas réparties de manière normale dans l'école.

9. Identification des élèves exceptionnels (par le ID ou index) :

Identifier les élèves qui se démarquent en calculant les écarts par rapport à la moyenne pour chaque caractéristique et en identifiant ceux qui ont des valeurs supérieures à un certain seuil (par exemple, les élèves dans le top 10% pour chaque caractéristique).

```
[87]: meilleur_22_23 = {}
for i in range(1, 10):
    # Calcul des écart avec la moyenne pour chaque catégorie:
    nomCatégorie = Année_22_23.columns[i] + "EcartMoy"
    Année_22_23[nomCatégorie] = Année_22_23[Année_22_23.columns[i]] - 5

    # Identifier les élèves qui ont des valeurs supérieur à un certain seuil:
    meilleurCategorie = Année_22_23[Année_22_23[Année_22_23.columns[i]] > 9]
    print("Il y a ", meilleurCategorie.shape[0], " élèves qui ont une note_
    ↳supérieur à 9 pour la caractéristique ", Année_22_23.columns[i], ".")
```

```
meilleur_22_23[Année_22_23.columns[i]] = meilleurCategorie
```

```
Il y a 20 élèves qui ont une note supérieur à 9 pour la caractéristique
Courage .
Il y a 0 élèves qui ont une note supérieur à 9 pour la caractéristique
Kindness .
Il y a 15 élèves qui ont une note supérieur à 9 pour la caractéristique
Lucidity .
Il y a 20 élèves qui ont une note supérieur à 9 pour la caractéristique
Intelligence .
Il y a 11 élèves qui ont une note supérieur à 9 pour la caractéristique
Creativity .
Il y a 16 élèves qui ont une note supérieur à 9 pour la caractéristique
Loyalty .
Il y a 15 élèves qui ont une note supérieur à 9 pour la caractéristique Cold
blood .
Il y a 19 élèves qui ont une note supérieur à 9 pour la caractéristique
Sincerity .
Il y a 8 élèves qui ont une note supérieur à 9 pour la caractéristique
Ambition .
```

- Pour identifier les élèves qui se démarquent dans chaque catégorie, j'ai choisi de prendre les élèves qui ont une note supérieur à 9 dans cette catégorie. J'ai ensuite mis ces dataset dans un dictionnaire dont la clé est la caractéristique. Ces dataset comportent aussi des colonnes avec l'écart avec la moyenne pour chaque caractéristique.

10. Dire, quelle maison possède les meilleurs élèves.

```
[88]: # Initialisation des moyennes
Gryffindor = 0
Hufflepuff = 0
Ravenclaw = 0
Slytherin = 0

# Addition des moyennes des caractéristiques par maisons
for i in range(1, 11):
    moyenne = Année_22_23.groupby('House')[Année_22_23.columns[i]].mean()
    Gryffindor += moyenne['Gryffindor']
    Hufflepuff += moyenne['Hufflepuff']
    Ravenclaw += moyenne['Ravenclaw']
    Slytherin += moyenne['Slytherin']

# Division par dix pour obtenir la moyenne des moyennes
Gryffindor = Gryffindor / 10
Hufflepuff = Hufflepuff / 10
Ravenclaw = Ravenclaw / 10
Slytherin = Slytherin / 10
```



```
# Affichage de la moyenne des maison + place dans le classement
print("Gryffindor:", Gryffindor, "(2)")
print("Hufflepuff:", Hufflepuff, "(1)")
print("Ravenclaw:", Ravenclaw, "(3)")
print("Slytherin:", Slytherin, "(4)")
```

```
Gryffindor: 5.145031413462422 (2)
Hufflepuff: 5.521689281301886 (1)
Ravenclaw: 5.050840932688221 (3)
Slytherin: 4.538215469667817 (4)
```

- Pour déterminer quelle maison à les meilleurs élèves, j'ai décider d'additionner la moyenne de chaque caractéristique par maison et puis de diviser cette somme par 10 faisant ainsi la moyenne de toutes les moyennes par maison.
- On voit que la maison avec les meilleurs élèves pour l'année 2022-2023 est Poufsoufle. Viens ensuite Gryffindor en deuxième position, Serdaigle en troisième et enfin Serpentard.

4 Partie 2 : Les qualités essentielles pour exceller en tant qu'élève (2023-2024)

Dans cette partie, vous répondez aux même questions que dans la partie 1 mais sur les données des élèves de 2023-2024.

0. ouvrir le dataset avec pandas et dire quelles sont les premières observations que tu fais sur ces données ?

```
[89]: Année_23_24 = pd.read_csv('hogwarts_23-24.csv', delimiter='\t')
      Année_23_24.head()
```

```
[89]:
```

	House	Courage	Kindness	Lucidity	Intelligence	Creativity	\
0	Slytherin	5.835231	2.074331	10.000000	2.107745	4.480949	
1	Slytherin	6.710366	4.142829	4.688952	3.689444	4.718914	
2	Slytherin	5.125860	2.571001	1.734415	3.260559	3.781189	
3	Slytherin	7.418516	3.273783	4.797849	5.813789	7.340648	
4	Slytherin	3.933961	2.567746	4.356123	3.350099	6.317330	

	Loyalty	Cold blood	Sincerity	Ambition	Wisdom
0	3.505679	4.343257	6.329798	2.511493	6.944614
1	4.246323	7.139023	5.560658	3.598693	3.906785
2	5.552486	5.923332	4.610046	4.118212	2.638779
3	2.239109	5.878091	4.380002	4.112518	3.528305
4	5.878184	8.512805	6.592798	4.743978	6.391331

- les élèves sont ranger dans l'ordre des maison. On à aussi beaucoup de caractéristique différentes sur les élèves.
1. Comment les élèves sont-ils répartis (en nombre) dans les différentes maisons (Gryffindor, Serpentard, Serdaigle, Poufsoufle) ?

```
[90]: maison = Année_23_24['House'].value_counts()
maison
```

```
[90]: House
Slytherin      200
Gryffindor     200
Ravenclaw      200
Hufflepuff     200
Name: count, dtype: int64
```

- Les élèves sont répartis de manière équitables entre chaque maison
2. Comment les élèves sont-ils répartis (en nombre) dans les différentes maisons (Gryffondor, Serpentard, Serdaigle, Poufsouffle) en fonction de leurs qualités pour le courage, la gentillesse et l'intelligence ?

```
[91]: Année_23_24['COURAGE'] = Année_23_24['Courage'] > 5
Courage = Année_23_24.groupby('House')['COURAGE'].value_counts()
print(Courage, "\n")

Année_23_24['GENTILLESSE'] = Année_23_24['Kindness'] > 5
Kindness = Année_23_24.groupby('House')['GENTILLESSE'].value_counts()
print(Kindness, "\n")

Année_23_24['INTELLIGENCE'] = Année_23_24['Intelligence'] > 5
Intelligence = Année_23_24.groupby('House')['INTELLIGENCE'].value_counts()
print(Intelligence, "\n")
```

House	COURAGE	
Gryffindor	False	111
	True	89
Hufflepuff	False	102
	True	98
Ravenclaw	True	146
	False	54
Slytherin	False	102
	True	98

Name: count, dtype: int64

House	GENTILLESSE	
Gryffindor	True	193
	False	7
Hufflepuff	True	168
	False	32
Ravenclaw	False	199
	True	1
Slytherin	False	197
	True	3

```
Name: count, dtype: int64
```

House	INTELLIGENCE	
Gryffindor	False	103
	True	97
Hufflepuff	False	131
	True	69
Ravenclaw	False	112
	True	88
Slytherin	False	120
	True	80

```
Name: count, dtype: int64
```

- COURAGE: Il y a de nombreux élèves courageux chez Serdaigle. Les élèves des autres maisons sont répartis de manière plutôt équitable. Cette qualité est devenue importante pour entrer chez Serdaigle mais n'a aucune influence sur l'appartenance aux autres maisons.
 - GENTILLESSE: On voit que les élèves de Gryffondor et de Poufsouffle sont presque tous au-dessus de la moyenne en terme de gentillesse. Inversement, presque tous les élèves de Serdaigle et de Serpentard sont en dessous de la moyenne. Cette qualité est donc importante dans la répartition des élèves dans les maisons.
 - INTELLIGENCE: Les élèves sont répartis de manière plutôt équitable entre les différentes maisons même si les élèves de Poufsouffle et de Serpentard sont en moyenne moins intelligents. Cette qualité ne doit pas être la plus importante dans la répartition des élèves dans les maisons.
3. Est-ce que les qualités (lucidité et intelligence) des élèves sont liées à leur appartenance (maison) ? Autrement, est-ce que le fait d'appartenir à une maison implique qu'on possède un peu plus ces 2 qualités ?

```
[92]: # On vérifie que la moyenne de ces deux qualités
# soit supérieure à la moyenne (5).
# Le "et" est matérialisé par le plus

Année_23_24['KindLucid'] = (Année_23_24['Intelligence'] +
↪Année_23_24['Lucidity']) / 2 > 5
kind_lucid = Année_23_24.groupby('House')['KindLucid'].value_counts()
kind_lucid
```

```
[92]: House      KindLucid
Gryffindor  True         111
           False        89
Hufflepuff  False       108
           True          92
Ravenclaw   False       101
           True          99
Slytherin   False       113
           True          87
```

Name: count, dtype: int64

- Les résultats sont plutôt omogènes donc ces qualité non sont pas importante pour la répartition des élèves dans les maisons.

4. Quelles sont les qualités les plus communes chez les élèves de Gryffondor ?

```
[93]: Com_quality = Année_23_24.groupby('House').mean()
      Com_quality.loc['Gryffindor'].sort_values() > 5
```

```
[93]: COURAGE          False
      INTELLIGENCE     False
      KindLucid        False
      GENTILLESSE      False
      Ambition         False
      Cold blood       False
      Courage          False
      Intelligence     False
      Creativity       False
      Loyalty          True
      Sincerity        True
      Lucidity         True
      Wisdom           True
      Kindness         True
      Name: Gryffindor, dtype: bool
```

- On voit que les qualité principales des élèves de Gryffindor sont: la gentillesse, la sagesse, la lucidité, la sincérité et la loyauté. Ces qualités ont changés depuis l'année dernière.

5. Moyennes des caractéristiques des élèves par maison :

Calculer la moyenne du courage, de l'intelligence, de la gentillesse, pour chaque maison (Gryffondor, Serpentard, Serdaigle, Poufsouffle) afin de déterminer les tendances centrales des caractéristiques des élèves dans chaque maison.

```
[94]: Moyenne = Année_23_24.groupby('House')[['Courage', 'Kindness', 'Intelligence']].
      ↪mean()
      Moyenne
```

```
[94]:
```

	Courage	Kindness	Intelligence
House			
Gryffindor	4.726738	6.752235	4.946895
Hufflepuff	5.125877	5.967494	4.458456
Ravenclaw	5.898829	3.184358	4.646639
Slytherin	4.892111	3.430651	4.477717

4.0.1 Tendances centrales des maisons:

- Gryffondor: Les élèves sont très **gentils** et ils ont une intelligence et un courage autour de la moyenne.
- Poufsouffle: Les élèves sont très **gentils** et ils ont un courage et une intelligence autour de la moyenne.
- Serdaigle: Les élèves ne sont pas très gentils. Ils sont très **courageux** et leur intelligence est dans la moyenne.
- Serpentard: Les élèves ont un courage et une intelligence autour de la moyenne mais ils ne sont pas gentils.

6. Moyenne générale des caractéristiques des élèves :

Calculer la moyenne globale du courage, de l'intelligence, de la gentillesse, pour tous les élèves afin de comprendre les tendances centrales des caractéristiques des élèves à l'échelle de l'école dans son ensemble.

```
[95]: Moy = Année_23_24[['Courage', 'Kindness', 'Intelligence']].mean()  
Moy
```

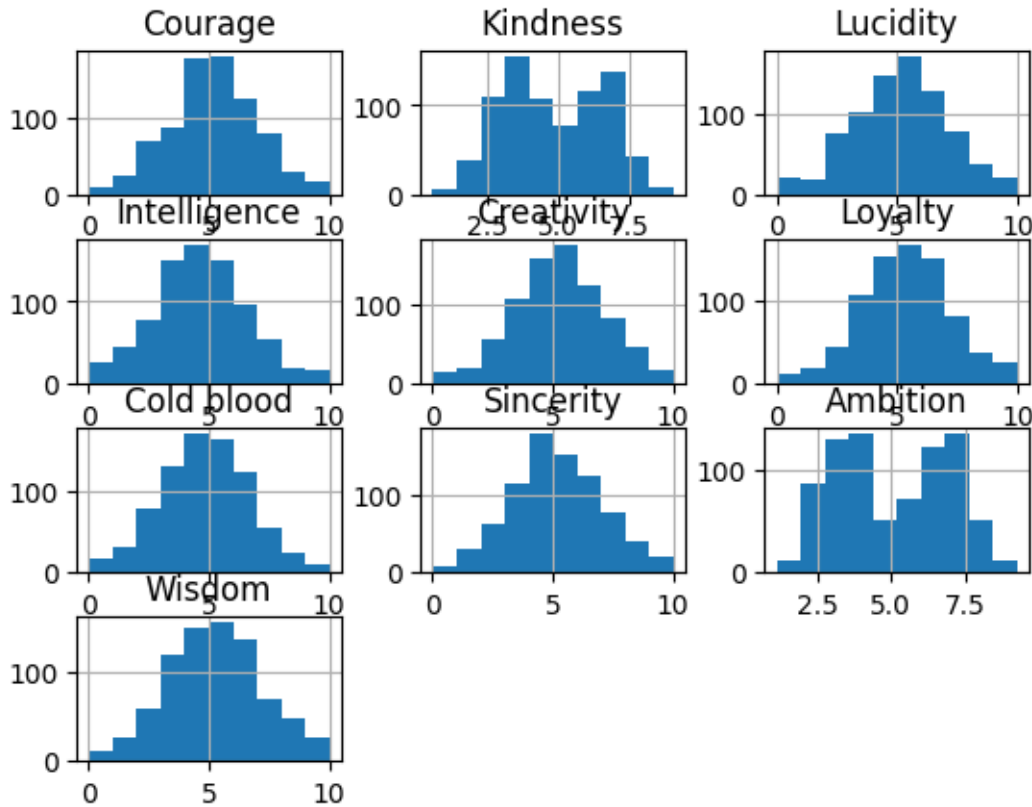
```
[95]: Courage          5.160889  
Kindness            4.833684  
Intelligence        4.632427  
dtype: float64
```

- Il n'y a pas de tendance spéciale qui se dégage particulièrement. On voit que la moyenne de l'école pour ces trois caractéristiques est autour de la moyenne.

7. Analyse de la distribution des caractéristiques :

Utiliser des histogrammes pour visualiser la distribution des caractéristiques des élèves (par exemple, le courage) et identifier si elles suivent une distribution normale ou non.

```
[96]: histo = Année_23_24[['Courage', 'Kindness', 'Lucidity', 'Intelligence',  
↪ 'Creativity', 'Loyalty', 'Cold blood', 'Sincerity', 'Ambition', 'Wisdom']].  
↪ hist()
```



- On voit que la plus part des caractéristiques sont répartis normalement: courage, lucidité, intelligence, créativité, loyal, sang-froid, sincérité, sagesse. Les graphiques épousent la courbe de Gauss.
- Inversement, la gentillesse et l'ambition ne suivent pas cette courbe. Ces deux caractéristiques ne sont pas réparties de manière normale dans l'école.

9. Identification des élèves exceptionnels (par le ID ou index) :

Identifier les élèves qui se démarquent en calculant les écarts par rapport à la moyenne pour chaque caractéristique et en identifiant ceux qui ont des valeurs supérieures à un certain seuil (par exemple, les élèves dans le top 10% pour chaque caractéristique).

```
[97]: meilleur_23_24 = {}
for i in range(1, 11):
    # Calcul des écart avec la moyenne pour chaque catégorie:
    nomCatégorie = Année_23_24.columns[i] + "EcartMoy"
    Année_23_24[nomCatégorie] = Année_23_24[Année_23_24.columns[i]] - 5

    # Identifier les élèves qui ont des valeurs supérieur à un certain seuil:
    meilleurCategorie = Année_23_24[Année_23_24[Année_23_24.columns[i]] > 9]
    print("Il y a ", meilleurCategorie.shape[0], " élèves qui ont une note_
    ↳supérieur à 9 pour la caractéristique ", Année_23_24.columns[i], ".")
```

```
meilleur_23_24[Année_23_24.columns[i]] = meilleurCategorie
```

```
Il y a 17 élèves qui ont une note supérieur à 9 pour la caractéristique
Courage .
Il y a 2 élèves qui ont une note supérieur à 9 pour la caractéristique
Kindness .
Il y a 22 élèves qui ont une note supérieur à 9 pour la caractéristique
Lucidity .
Il y a 16 élèves qui ont une note supérieur à 9 pour la caractéristique
Intelligence .
Il y a 18 élèves qui ont une note supérieur à 9 pour la caractéristique
Creativity .
Il y a 25 élèves qui ont une note supérieur à 9 pour la caractéristique
Loyalty .
Il y a 9 élèves qui ont une note supérieur à 9 pour la caractéristique Cold
blood .
Il y a 20 élèves qui ont une note supérieur à 9 pour la caractéristique
Sincerity .
Il y a 4 élèves qui ont une note supérieur à 9 pour la caractéristique
Ambition .
Il y a 26 élèves qui ont une note supérieur à 9 pour la caractéristique
Wisdom .
```

- Pour identifier les élèves qui se démarquent dans chaque catégorie, j'ai choisi de prendre les élèves qui ont une note supérieure à 9 dans cette catégorie. J'ai ensuite mis ces données dans un dictionnaire dont la clé est la caractéristique. Ces données comportent aussi des colonnes avec l'écart avec la moyenne pour chaque caractéristique.

10. Dire, quelle maison possède les meilleurs élèves.

```
[98]: # Initialisation des moyennes
Gryffindor = 0
Hufflepuff = 0
Ravenclaw = 0
Slytherin = 0

# Addition des moyennes des caractéristiques par maisons
for i in range(1, 11):
    moyenne = Année_23_24.groupby('House')[Année_23_24.columns[i]].mean()
    Gryffindor += moyenne['Gryffindor']
    Hufflepuff += moyenne['Hufflepuff']
    Ravenclaw += moyenne['Ravenclaw']
    Slytherin += moyenne['Slytherin']

# Division par dix pour obtenir la moyenne des moyennes
Gryffindor = Gryffindor / 10
Hufflepuff = Hufflepuff / 10
Ravenclaw = Ravenclaw / 10
```

```
Slytherin = Slytherin / 10

# Affichage de la moyenne des maison + place dans le classement
print("Gryffindor:", Gryffindor, "(3)")
print("Hufflepuff:", Hufflepuff, "(1)")
print("Ravenclaw:", Ravenclaw, "(2)")
print("Slytherin:", Slytherin, "(4)")
```

```
Gryffindor: 5.009506092710777 (3)
Hufflepuff: 5.53780118538631 (1)
Ravenclaw: 5.081716646569703 (2)
Slytherin: 4.657745473418089 (4)
```

- Pour déterminer quelle maison a les meilleurs élèves, j'ai décidé d'additionner la moyenne de chaque caractéristique par maison et puis de diviser cette somme par 10 faisant ainsi la moyenne de toutes les moyennes par maison.
- On voit que la maison avec les meilleurs élèves pour l'année 2023-2024 est à nouveau Poufsouffle. Viens ensuite Serdaigne en deuxième position, Gryffindor en troisième et enfin Serpentard encore dernière. Gryffindor et Serdaigne ont échangé leurs positions.

5 Partie 3 : Analysez

Dans cette partie, vous avez 8 questions, et vous devez faire preuve d'imagination pertinente en ajoutant 2 questions au choix (venant de vous, bien évidemment :)) .

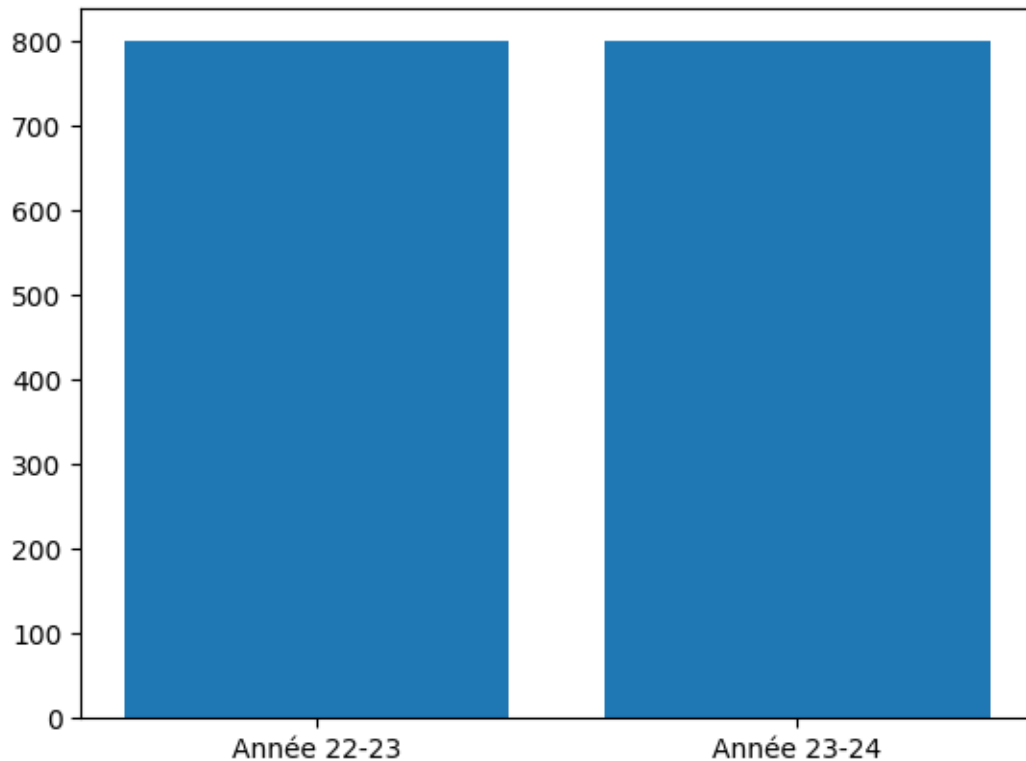
Vous allez maintenant observer, analyser (avec des visualisations graphiques à l'appui), les deux années et ce que l'on attend de vous c'est que vous soyez en mesure de conclure si oui ou non vous observez une amélioration des qualités entre les deux années.

1. Quelle est la différence de taille entre les deux jeux de données en termes de nombre d'élèves ?

```
[99]: print(Année_22_23.shape)
      print(Année_23_24.shape)

plt.bar(height=[Année_22_23.shape[0], Année_23_24.shape[0]], x=['Année 22-23',
↪ 'Année 23-24'])
plt.show()
```

```
(800, 24)
(800, 25)
```

- La taille des deux jeux de donnée est la même. Il y a le même nombre d'élève et le même nombre de caractéristique.
2. Y a-t-il des différences dans la répartition des élèves entre les maisons (Gryffondor, Serpentard, Serdaigle, Poufsouffle) d'une année à l'autre ?

```
[100]: a = Année_22_23['House'].value_counts()
b = Année_23_24['House'].value_counts()

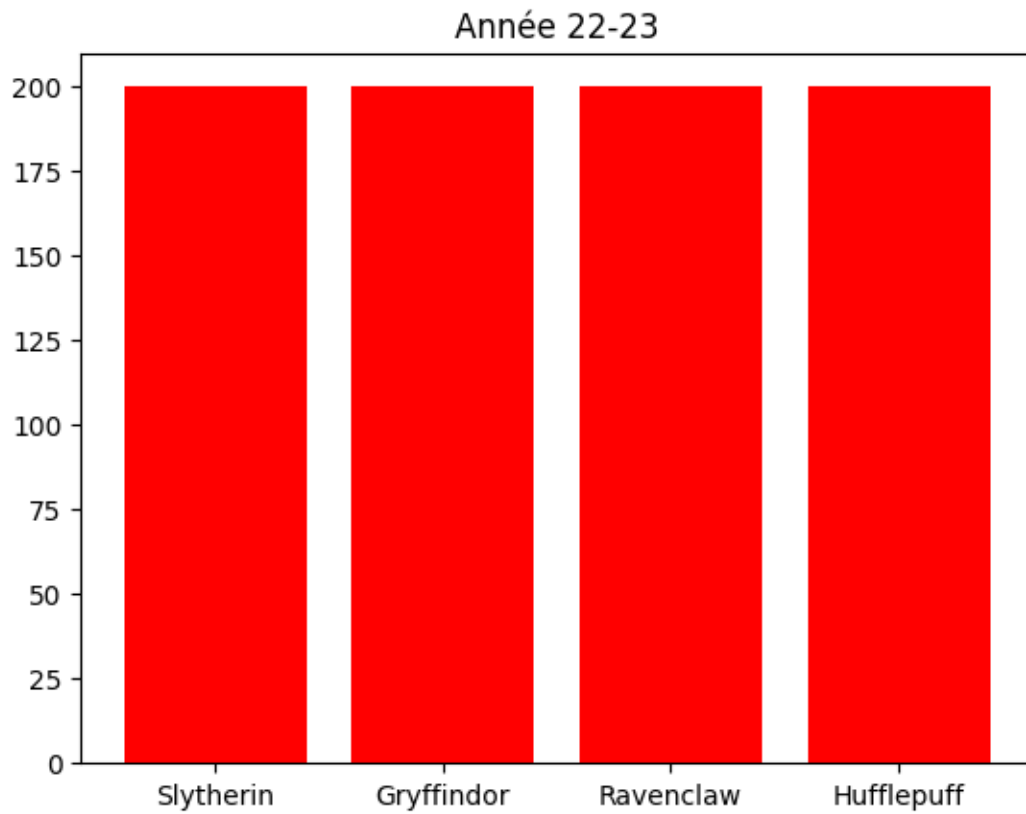
print("Année 22-23")
print(a)
print()
print("Année 23-24")
print(b)
print()

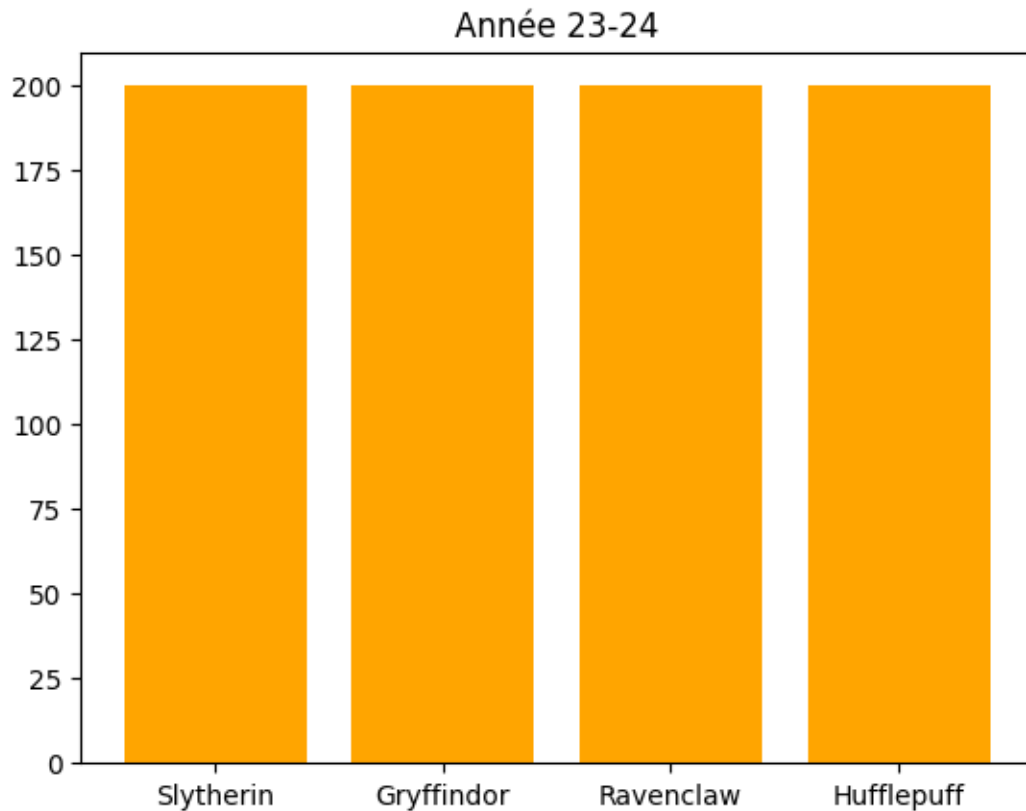
plt.bar(height=a, x=a.index, color="red")
plt.title("Année 22-23")
plt.show()

plt.bar(height=b, x=b.index, color="orange")
plt.title("Année 23-24")
plt.show()
```

```
Année 22-23
House
Slytherin      200
Gryffindor     200
Ravenclaw      200
Hufflepuff     200
Name: count, dtype: int64
```

```
Année 23-24
House
Slytherin      200
Gryffindor     200
Ravenclaw      200
Hufflepuff     200
Name: count, dtype: int64
```





- Il n'y a pas de différence dans la répartition des élèves dans les maison. Elles ont toutes 200 élèves.
3. Quelles sont les moyennes des caractéristiques des élèves (courage, gentillesse, intelligence, etc.) pour chaque année et comment ont-elles évolué ?

```
[101]: def tendance_variation_caracteristique(caracteristique):
    Caracteristique_22_23 = Année_22_23[caracteristique].mean()
    print("Année 22-23")
    print(caracteristique, ":", Caracteristique_22_23)
    print()

    Caracteristique_23_24 = Année_23_24[caracteristique].mean()
    print("Année 23-24")
    print(caracteristique, ":", Caracteristique_23_24)
    print()

    plt.plot([22, 24], [Caracteristique_22_23, Caracteristique_23_24])
    plt.title("Evolution moyenne " + caracteristique)
    plt.show()

for i in range(1, 11):
```

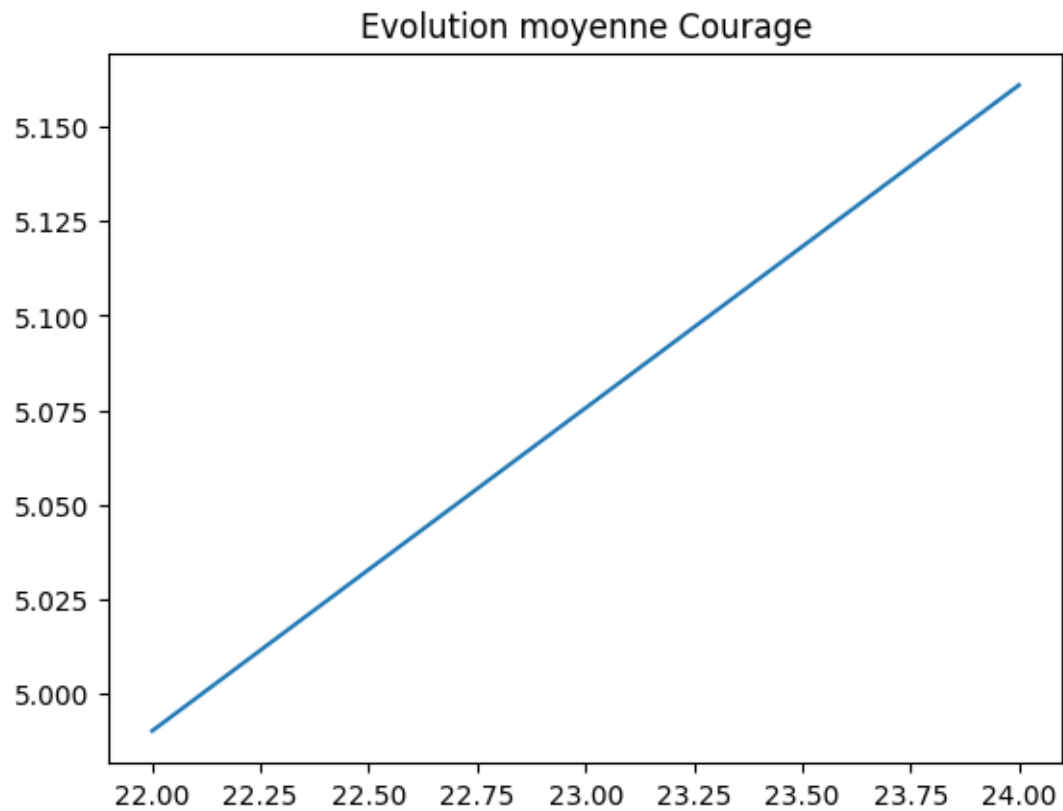
```
tendance_variation_caracteristique(Année_22_23.columns[i])
```

Année 22-23

Courage : 4.990261456185708

Année 23-24

Courage : 5.160889068049592

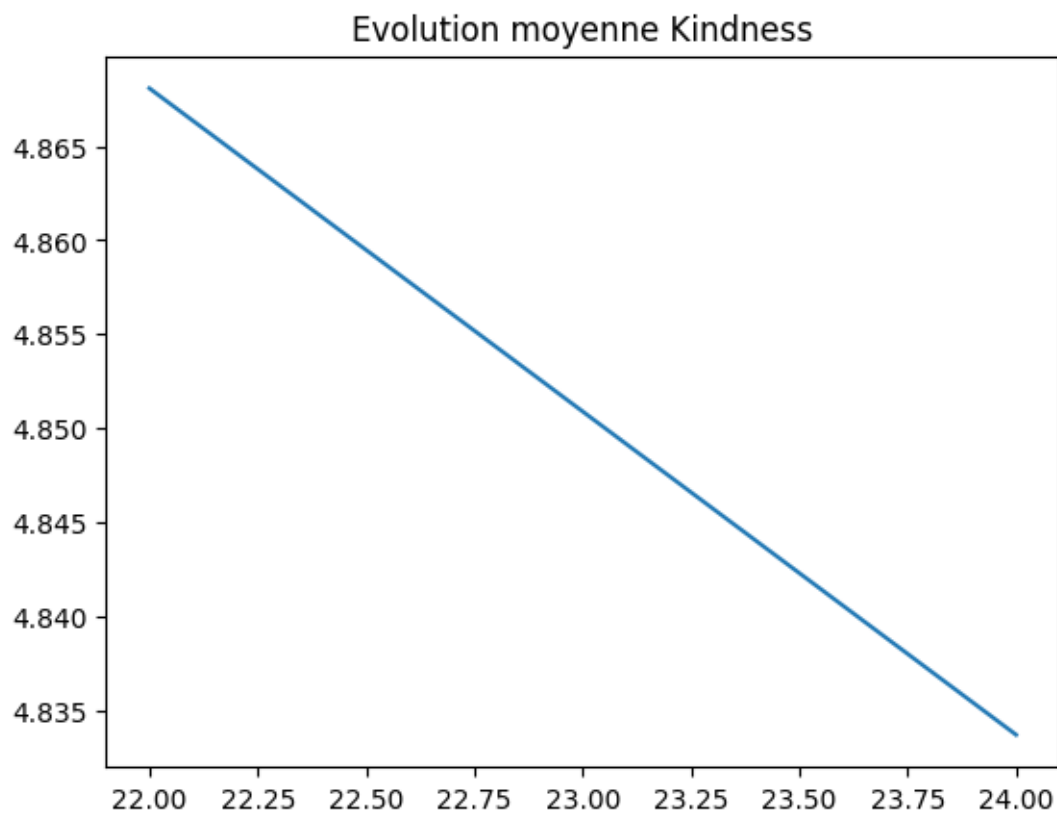


Année 22-23

Kindness : 4.8680756113332295

Année 23-24

Kindness : 4.833684286459194

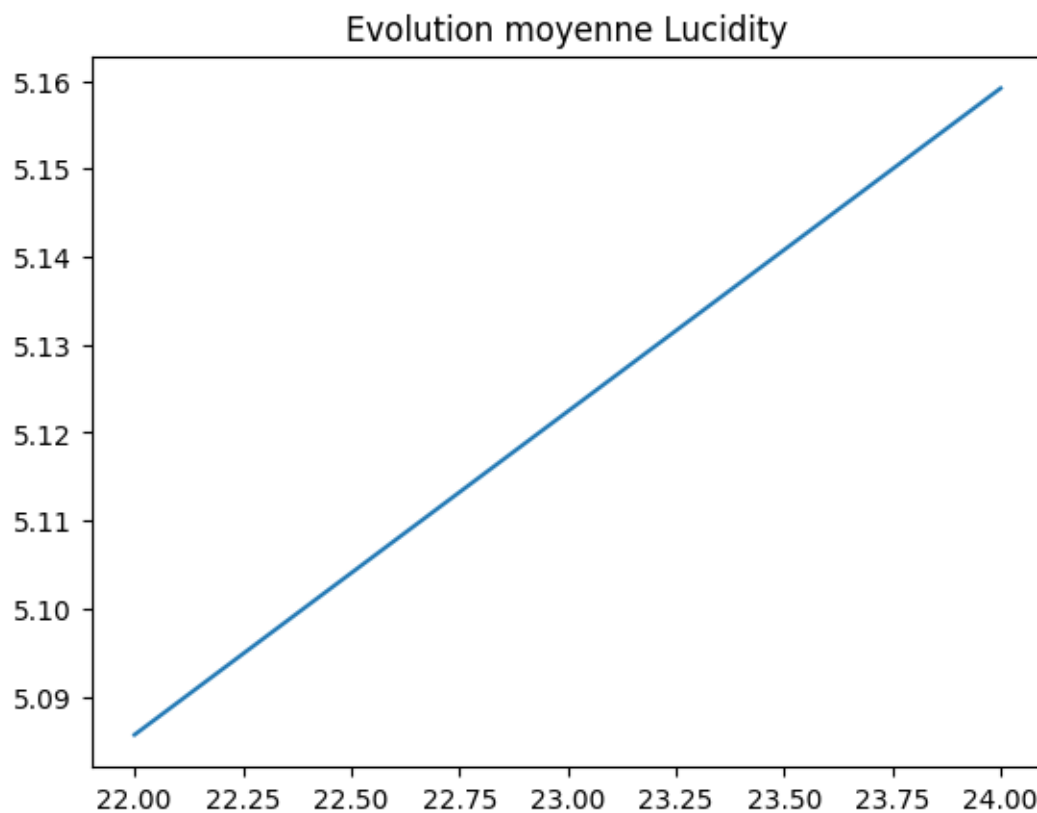


Année 22-23

Lucidity : 5.085657507113255

Année 23-24

Lucidity : 5.159130885598795

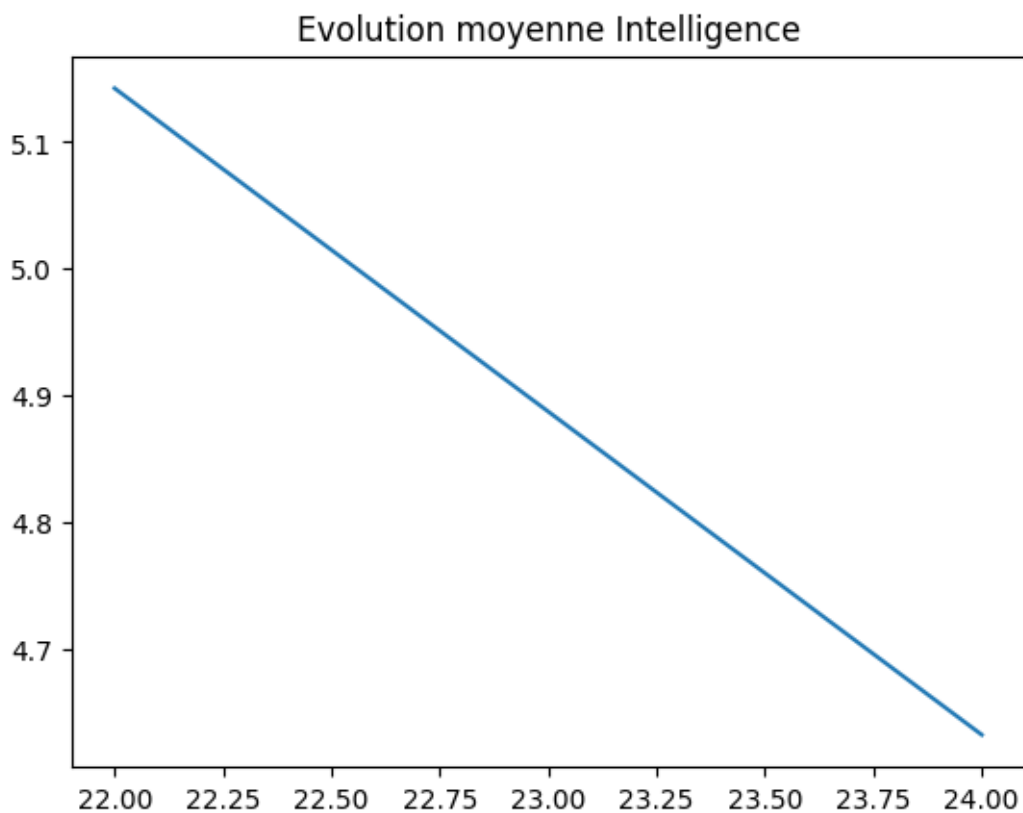


Année 22-23

Intelligence : 5.1422950347289955

Année 23-24

Intelligence : 4.632426640594257

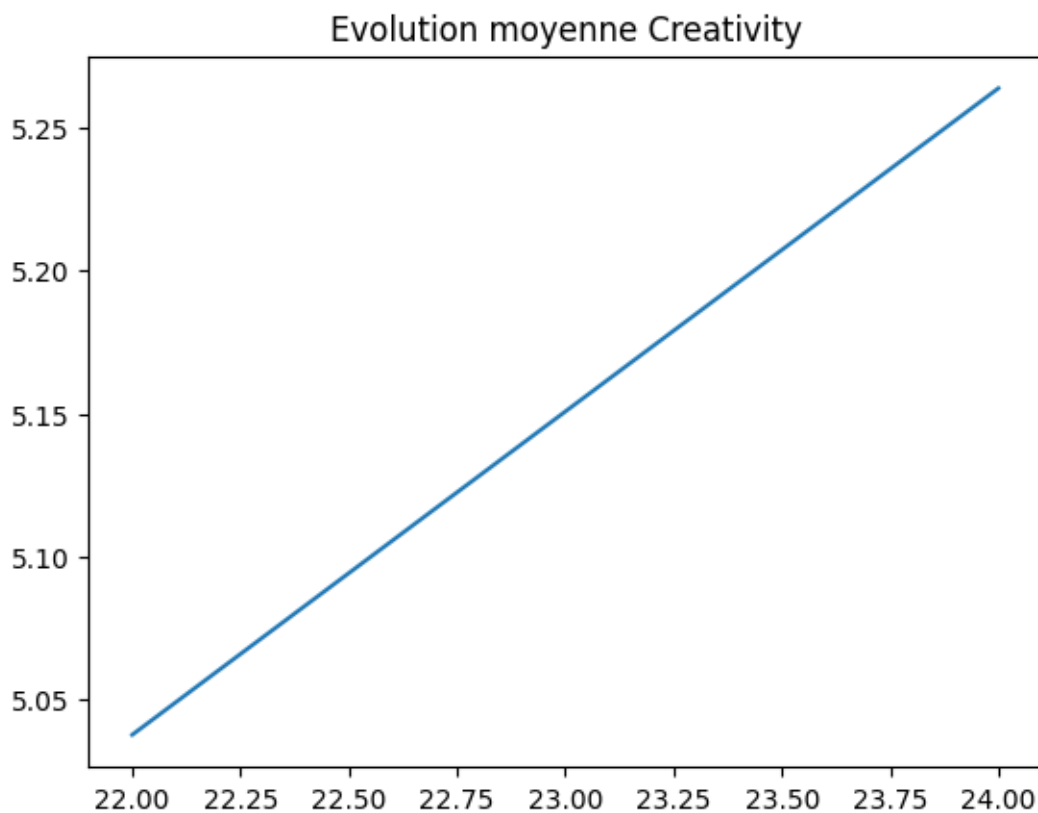


Année 22-23

Creativity : 5.03774259144401

Année 23-24

Creativity : 5.263699623370183

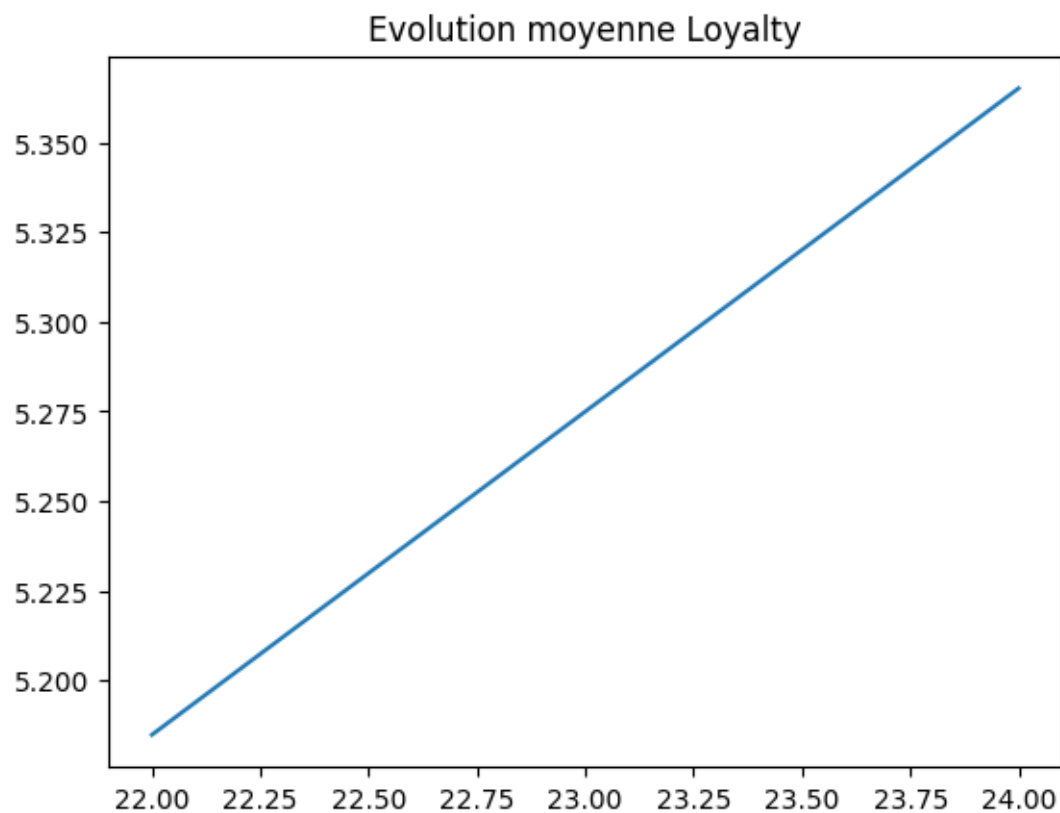


Année 22-23

Loyalty : 5.184807703263839

Année 23-24

Loyalty : 5.365128117735157

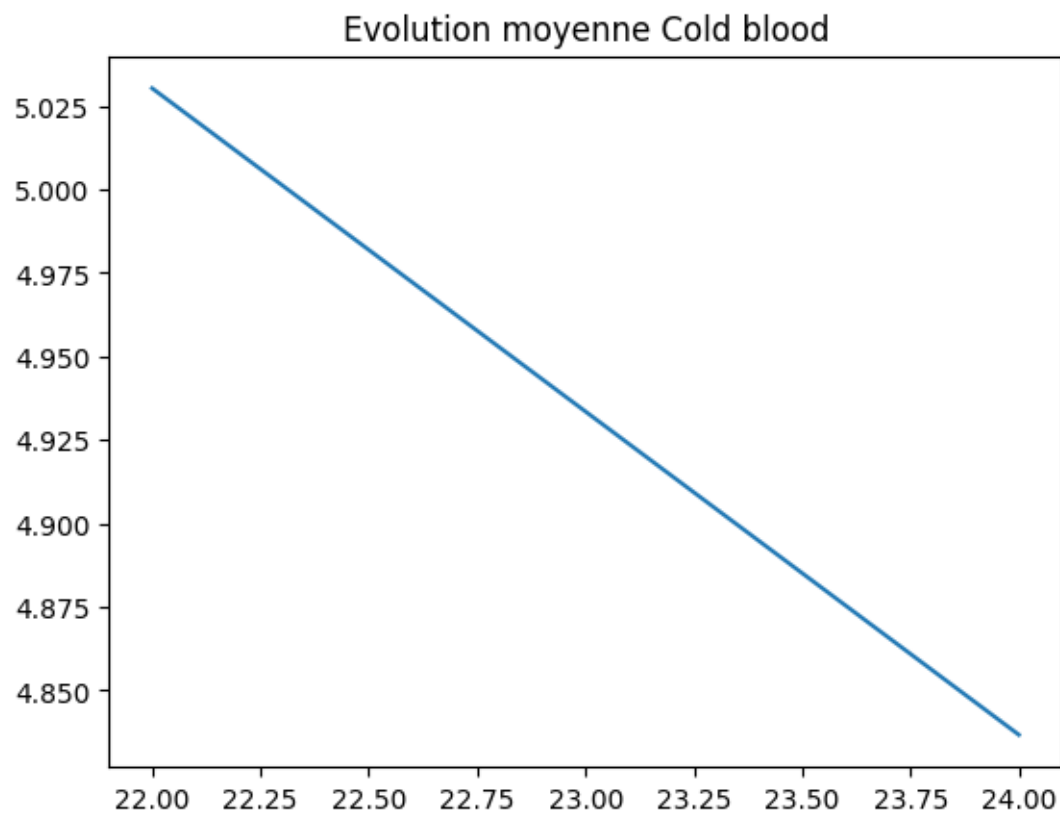


Année 22-23

Cold blood : 5.030213749962542

Année 23-24

Cold blood : 4.836755532502903

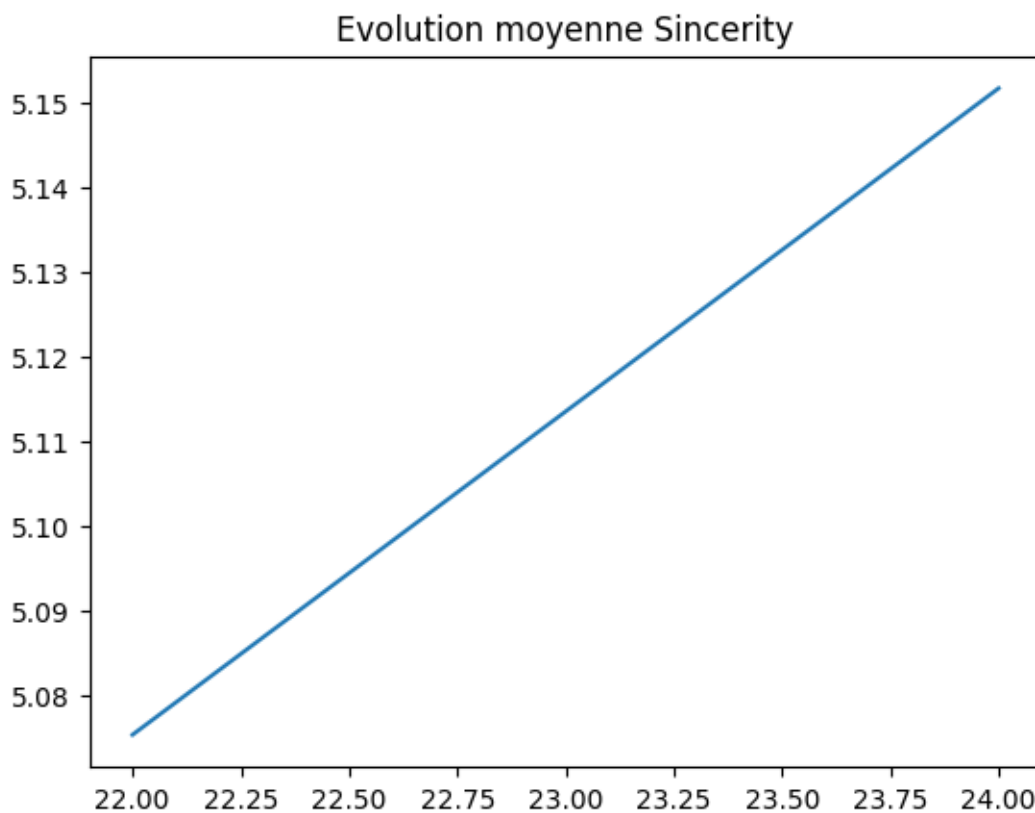


Année 22-23

Sincerity : 5.075358990576217

Année 23-24

Sincerity : 5.1517440317352365

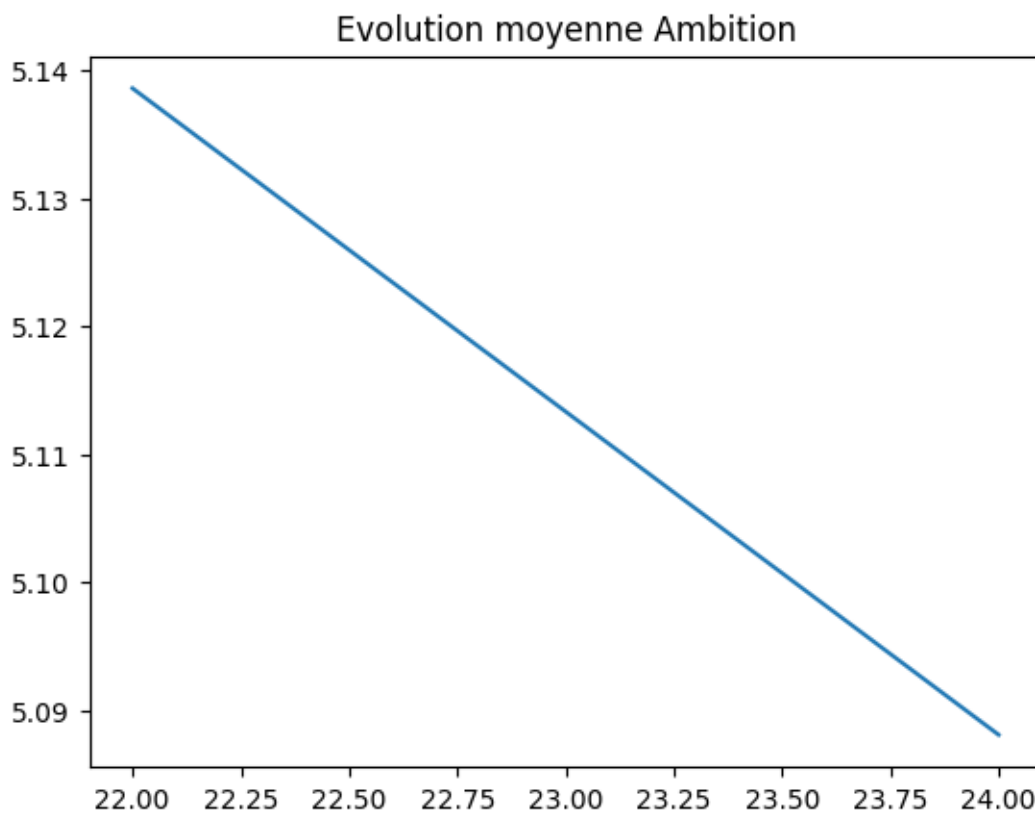


Année 22-23

Ambition : 5.13857204525518

Année 23-24

Ambition : 5.088078152366412

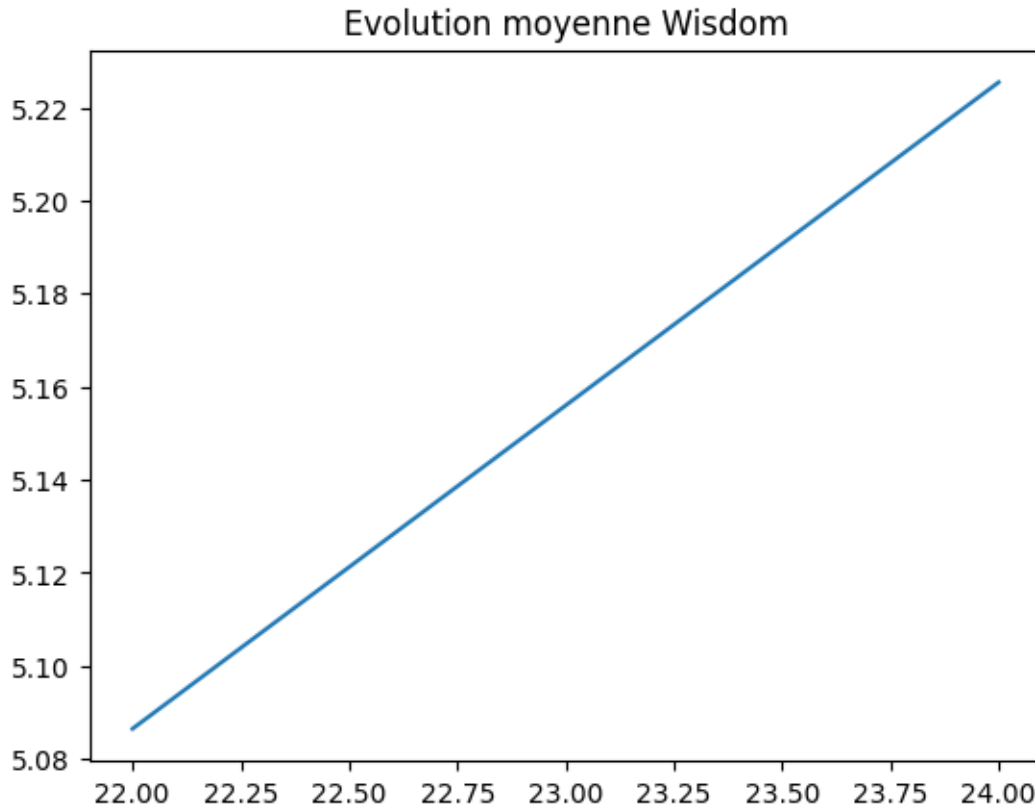


Année 22-23

Wisdom : 5.086458052937896

Année 23-24

Wisdom : 5.225387156800467



- Voici les caractéristique qui ont évoluer à la hausse: Windsom, Sincerity, Loyalty, Creativity, Lucidity et Courage.
 - Voici les caractéristique qui ont évoluer à la baisse: Ambition, Cold Blood, Intelligence, Kindness.
4. Existe-t-il des tendances dans la variation des caractéristiques des élèves d'une année à l'autre ?
- cf graphique Q3
 - Les tendances dans la variation des caractéristique des élèves d'une années à l'autre est en lien avec l'évolution haussière ou baissière de la moyenne de la caractéristique (c'est logique ...):
 - Voici les caractéristique qui ont une tendance haussière: Windsom, Sincerity, Loyalty, Creativity, Lucidity et Courage.
 - Voici les caractéristique qui ont une tendance baissière: Ambition, Cold Blood, Intelligence, Kindness.
5. Comment la répartition des élèves par maison a-t-elle changé d'une année à l'autre en fonction de leurs caractéristiques ?

```
[102]: def question5(maison):
        print(maison + ":")

        Maison_22_23 = Année_22_23.loc[Année_22_23['House'] == maison]
        print("Année 22-23")
        print(Maison_22_23[['Courage', 'Kindness', 'Lucidity', 'Intelligence',
        ↳ 'Creativity', 'Loyalty', 'Cold blood', 'Sincerity', 'Ambition', 'Wisdom']]).
        ↳ hist()
        plt.show()

        Maison_23_24 = Année_23_24.loc[Année_23_24['House'] == maison]
        print("Année 23-24")
        print(Maison_23_24[['Courage', 'Kindness', 'Lucidity', 'Intelligence',
        ↳ 'Creativity', 'Loyalty', 'Cold blood', 'Sincerity', 'Ambition', 'Wisdom']]).
        ↳ hist()

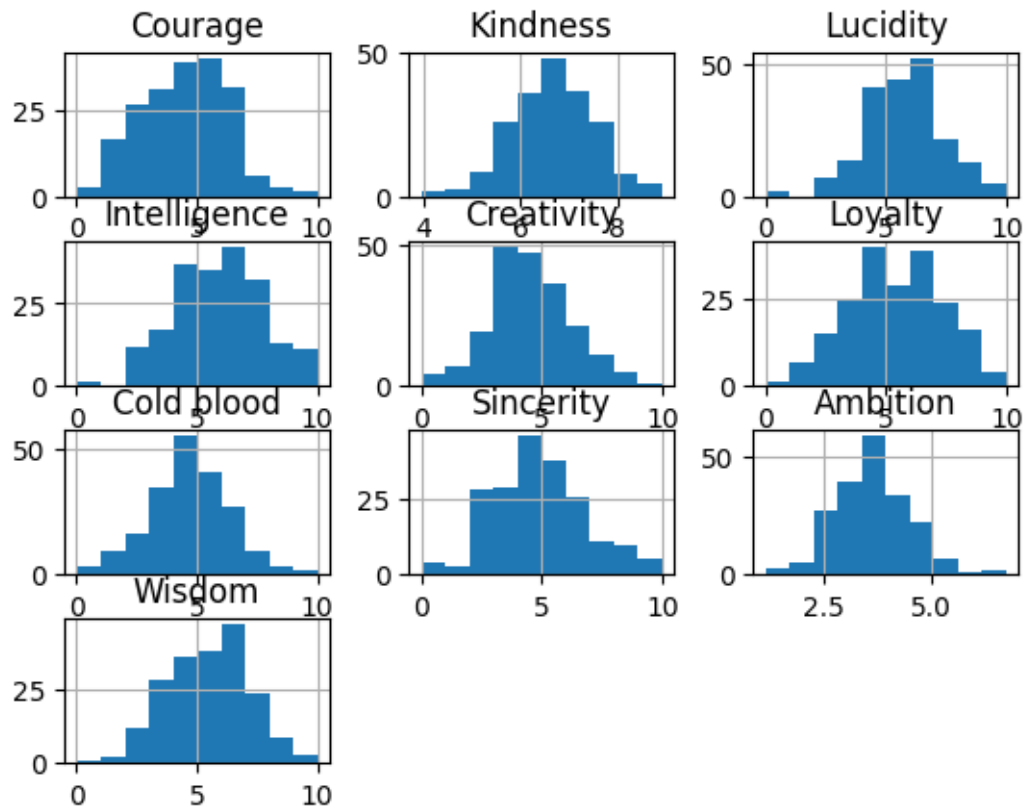
        ↳
        ↳ print("-----")
        ↳ plt.show()

question5("Gryffindor")
question5("Hufflepuff")
question5("Ravenclaw")
question5("Slytherin")
```

Gryffindor:

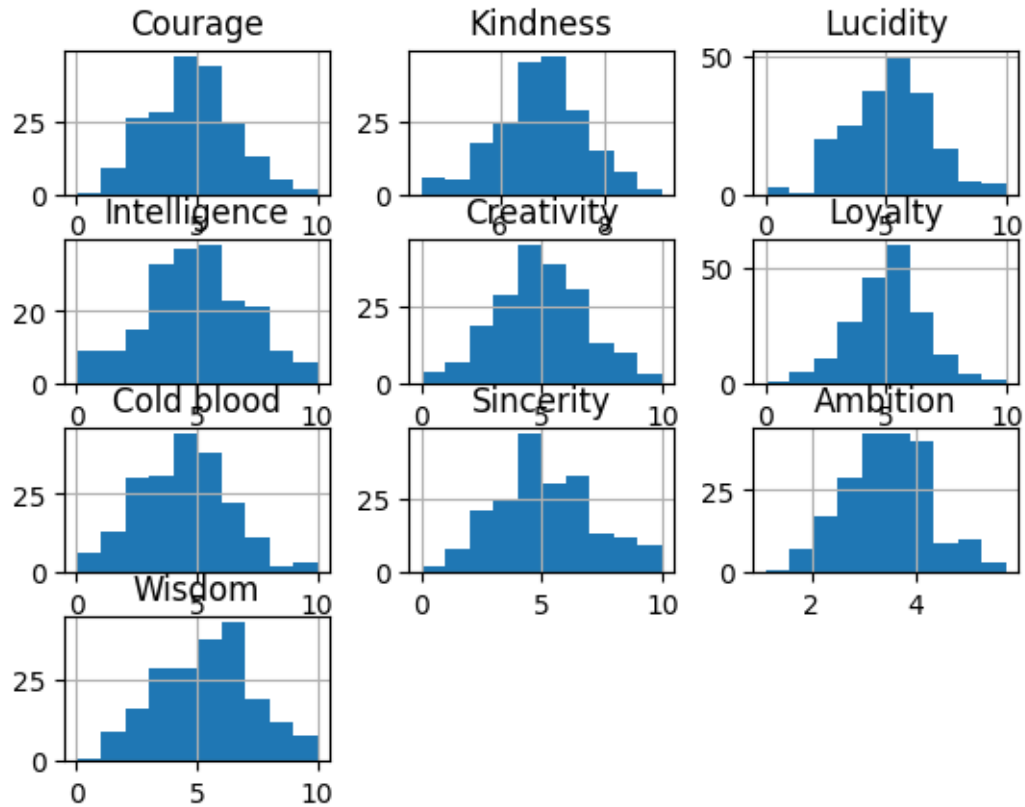
Année 22-23

```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>]
[(<Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>]
[(<Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>]
[(<Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Année 23-24

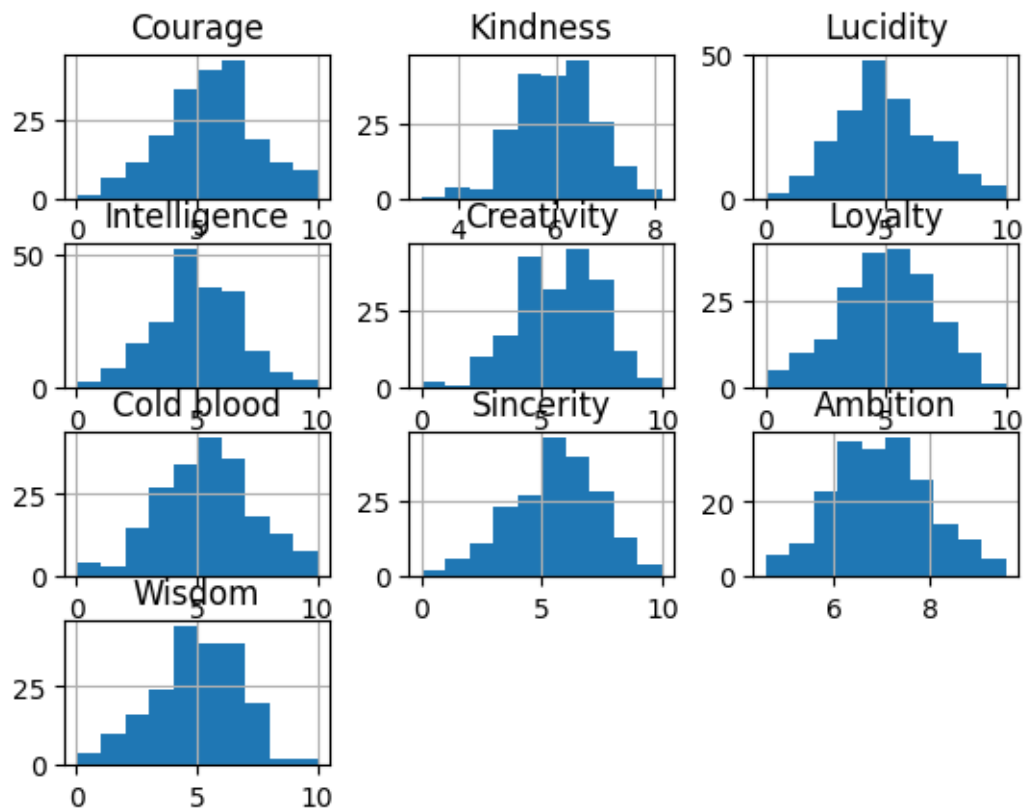
```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
  <Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
  <Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
  <Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Hufflepuff:

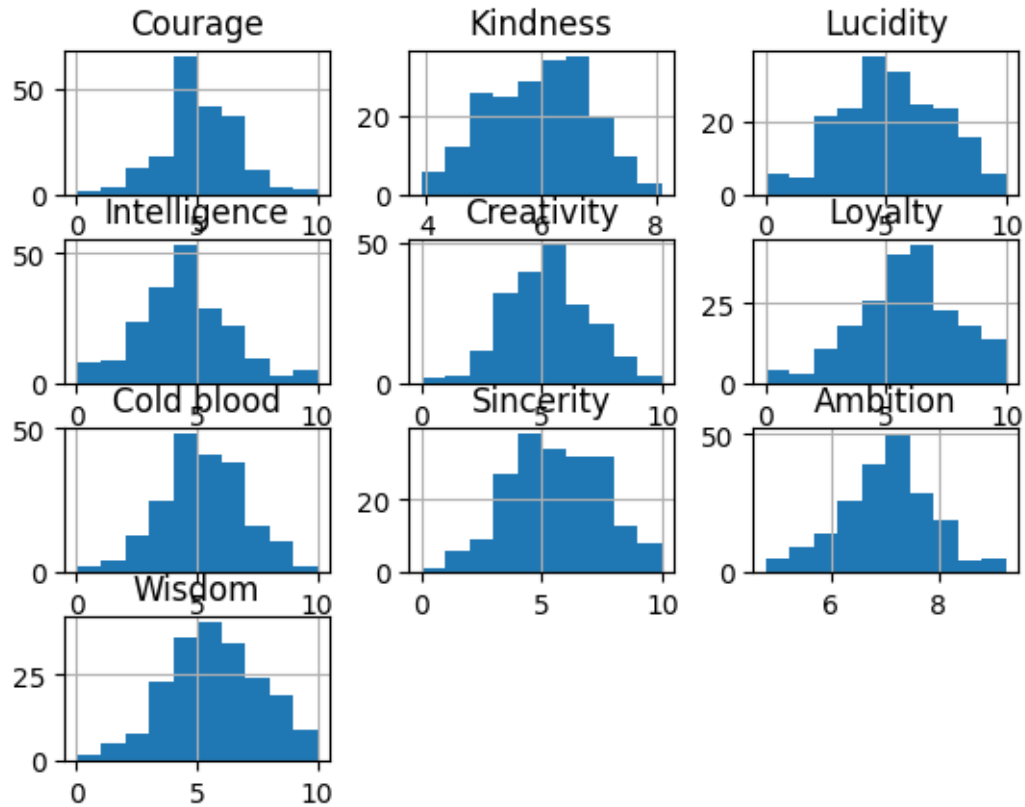
Année 22-23

```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
<Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
<Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
<Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```

Année 23-24

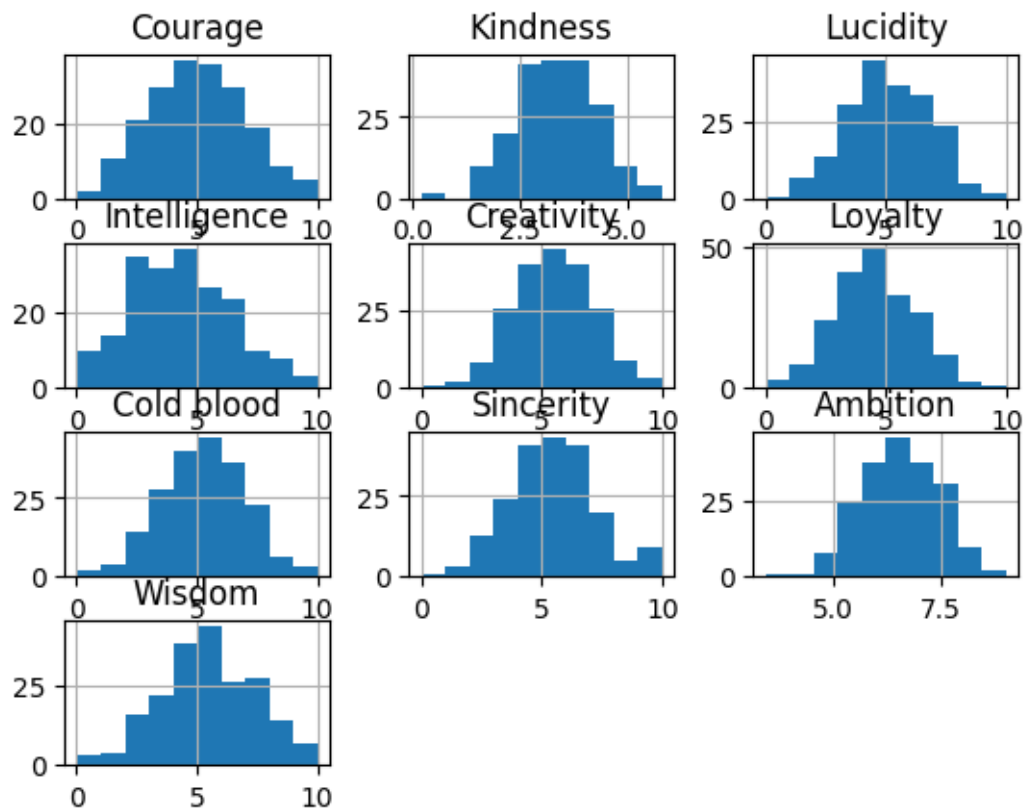
```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
  <Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
  <Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
  <Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Ravenclaw:

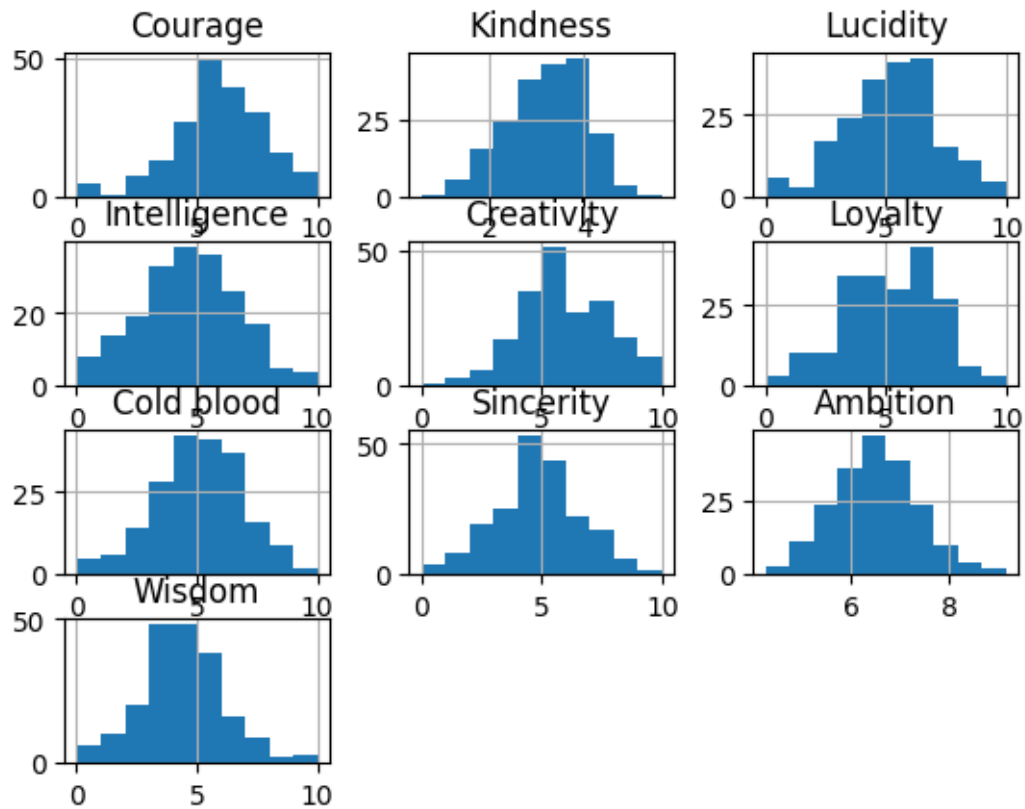
Année 22-23

```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>]
[(<Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>]
[(<Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>]
[(<Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Année 23-24

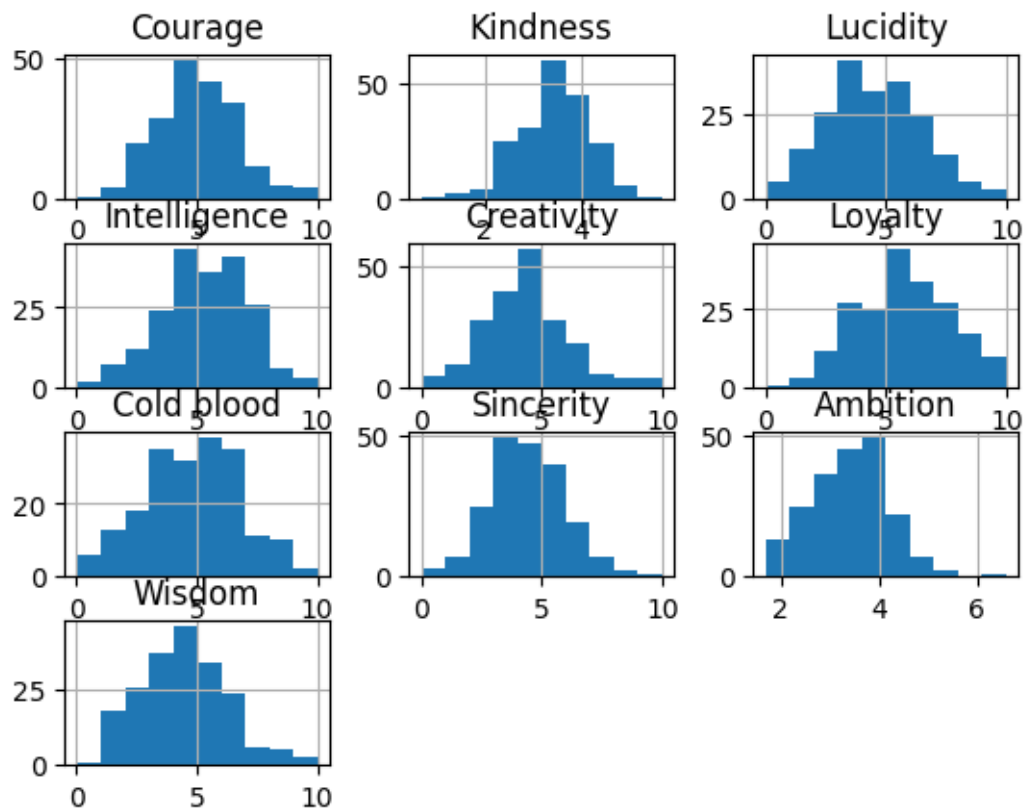
```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
<Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
<Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
<Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Slytherin:

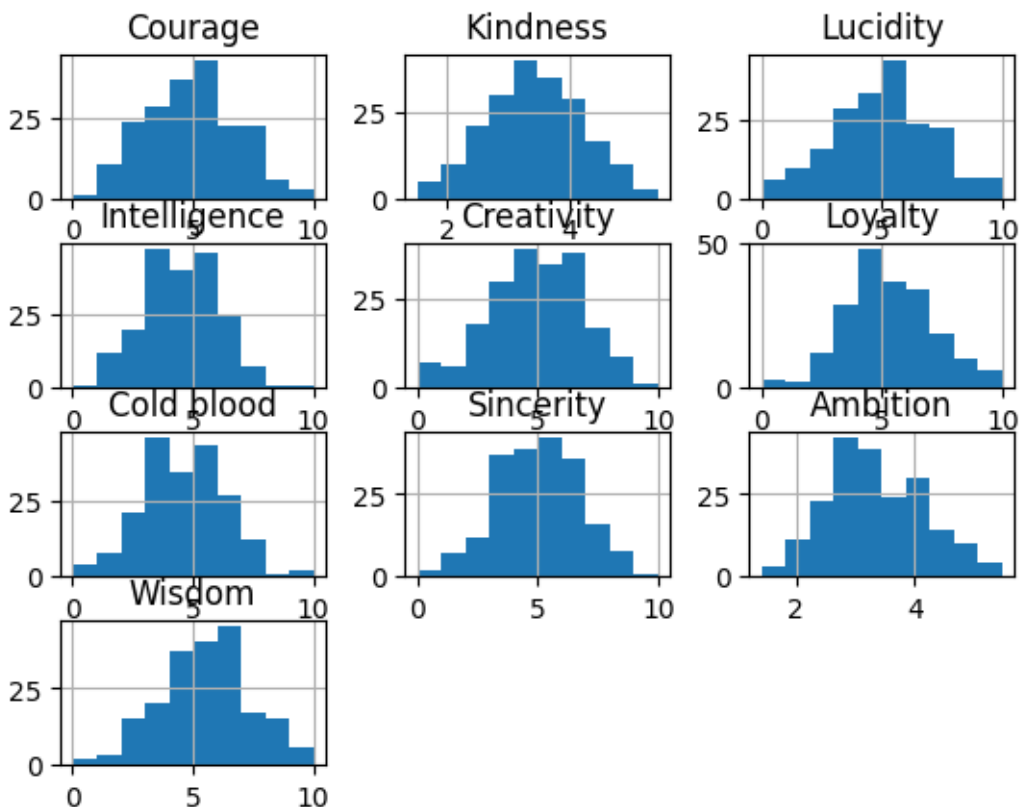
Année 22-23

```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
  <Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
  <Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
  <Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



Année 23-24

```
[(<Axes: title={'center': 'Courage'}>
  <Axes: title={'center': 'Kindness'}>
  <Axes: title={'center': 'Lucidity'}>
  <Axes: title={'center': 'Intelligence'}>
  <Axes: title={'center': 'Creativity'}>
  <Axes: title={'center': 'Loyalty'}>
  <Axes: title={'center': 'Cold blood'}>
  <Axes: title={'center': 'Sincerity'}>
  <Axes: title={'center': 'Ambition'}>
  <Axes: title={'center': 'Wisdom'}> <Axes: > <Axes: >)]
```



- Voici comment la répartition des élèves par maison a-t-elle changé d'une année à l'autre en fonction de leurs caractéristiques:
 - Gryffondor:
 - * Les élèves sont plus courageux, on plus de sang froid et ils ont plus d'ambition la deuxième année.
 - Poufsouffle:
 - * la répartition des élèves est plus omogène sur l'éventail de note en deuxième année.
 - Serdaigne:
 - * En moyenne, les élèves ont tous évoluer à la hausse entre les deux année.
 - Serpentard:
 - * · En moyenne, les élèves ont tous évoluer à la baisse entre les deux année.

6. Y a-t-il des similitudes ou des différences remarquables dans la distribution des caractéristiques des élèves entre les deux années ?

[]:

7. Quelle est la variation de la performance académique moyenne des élèves entre les deux années et quelles sont les caractéristiques qui y sont associées ?

[]:

8. Quelles sont les principales conclusions que l'on peut tirer de la comparaison entre les deux jeux de données ?

[]:

9.
10.

6 Conclusion

Bravo Harry/Hermione , vous avez relevé le défis , ce sera donc 10 points pour Griffondor .

Maintenant il est temps pour vous de soumettre ce projet, avant le 24 février (c.à.d. au plus tard le 23 février à 23H59 sur moodle), avec un simple « Wingardium Leviosa » .