

# Les Jeux Olympiques

## Table des matières

I.	Présentation du projet .....	1
II.	Introduction .....	1
III.	Analyse.....	1
1.	Etude statistique univariée des variables.....	2
2.	Etude de l'évolution du tableau des médailles à travers le temps .....	9
	Les Etats-Unis.....	9
	La Russie .....	10
3.	Etude des performances des équipes en fonction de si elles sont à domicile ou non .....	12
	Analyse brute des données .....	12
	Nuancement .....	13
	Explications des résultats de l'analyse .....	14
4.	Etude des performances des athlètes en fonction leur âge et de leur sexe.....	15
	Âge moyen des athlètes médaillés :.....	15
	Réponse à la problématique .....	18
IV.	Conclusion.....	18

### I. Présentation du projet

Les Jeux Olympiques de Paris 2024 approchent à grands pas et nous sommes en droit de se demander quels facteurs pourraient jouer sur les performances de nos athlètes français durant ces olympiades. En effet vous l'aurez compris nous avons choisi comme thème les Jeux Olympiques, étant donné que les Jeux Olympiques d'été ont lieu cette année à Paris, nous nous sommes sentis touchés par ce sujet.

### II. Introduction

Nous avons donc procédé à une étude générale des données. Ensuite nous avons tenté de répondre à trois problématiques. En premier nous avons étudié l'évolution du tableau des médailles à travers le temps afin d'en apprendre plus sur les évolutions remarquables de certains pays. Puis nous nous sommes attardés sur l'étude des performances des équipes en fonction de si elles sont à domicile ou non, ce qui nous permet de répondre à la question : a-t-on des chances en 2024 de gagner plus de médailles que lors des précédentes olympiades. Enfin, nous avons effectué une étude sur les performances des athlètes olympiques en fonction de leur âge et de leur sexe.

### III. Analyse

## 1. Etude statistique univariée des variables

Pour procéder à une étude univariée des variables nous allons tout d'abord chercher à déterminer quels types d'informations apportent ce DataFrame nommée JO, grâce à l'utilisation du logiciel python.

On affiche les 5 premières variables du jeu de donnée :

Out[5]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

On remarque que le jeu de données est composé de 15 variables différentes : le numéro d'identifiant ID du participant, son nom et prénom Name, son sexe Sex, son age Age, sa taille Height, son poids Weight, le pays que le participant représente Team, le code du pays NOC, le titre des jeux olympiques Games, l'année des jeux olympiques Year, sa saison Season, la ville où se passe les jeux olympiques City, la spécialité sportive du participant Sport, le nom de la compétition Event, et la médaille gagné par le participant Medal.

En étudiant sur python le jeu de données, on découvre qu'il y a 271 116 individus.

Vérifions si des variables ont des données manquantes :

```
Data columns (total 15 columns):
#   Column      Non-Null Count  Dtype
---  -
0    ID           271116 non-null   int64
1    Name         271116 non-null   object
2    Sex          271116 non-null   object
3    Age          261642 non-null   float64
4    Height       210945 non-null   float64
5    Weight       208241 non-null   float64
6    Team         271116 non-null   object
7    NOC          271116 non-null   object
8    Games        271116 non-null   object
9    Year         271116 non-null   int64
10   Season       271116 non-null   object
11   City         271116 non-null   object
12   Sport        271116 non-null   object
13   Event        271116 non-null   object
14   Medal        39783 non-null    object
dtypes: float64(3), int64(2), object(10)
memory usage: 31.0+ MB
```

On observe sur le tableau ci-dessus que les variables non-nulles ne sont pas toutes au nombre de 271 116, signifiant que certaines variables ont des valeurs nulles. Ces variables sont : Age, Height, Weight et Medal.

Nous avons complété ces informations manquantes (cf. annexe)

Voilà par exemple les données du 13e individu :

Et les données de la 4e variable :

```

Les données du 13eme individu sont :
ID                               6
Name                             Per Knut Aaland
Sex                               M
Age                              31.0
Height                           188.0
Weight                           75.0
Team                             United States
NOC                              USA
Games                           1992 Winter
Year                             1992
Season                           Winter
City                             Albertville
Sport                           Cross Country Skiing
Event                           Cross Country Skiing Men's 4 x 10 kilometres R...
Medal                            NaN
Name: 13, dtype: object

Les données de la 4eme variable sont :
0      180.0
1      170.0
2         NaN
3         NaN
4      185.0
...
271111  179.0
271112  176.0
271113  176.0
271114  185.0
271115  185.0
Name: Height, Length: 271116, dtype: float64

```

Le tableau suivant indique le type de chaque variable, où Object désigne un ou plusieurs caractères :

Variable	Type
ID	int64
Name	object
Sex	object
Age	float64
Height	float64
Weight	float64
Team	object
NOC	object
Games	object
Year	int64
Season	object
City	object
Sport	object
Event	object
Medal	object

On remarque alors que les seules variables qui ne sont pas des caractères sont ID et Year, qui sont des nombres entiers (Int), et Height et Weight sont des nombres réels.

Déterminons les modalités de la variable Sex et le nombre de fois où chacune apparait, avec la commande :

```

Sex
M    196594
F     74522
Name: count, dtype: int64

```

Il y a donc 196 594 participants de sexe masculin, et 74522 de sexe féminin. Intéressons-nous aux variables Name, Age, Height, Team et Games :

Name :

Age :

Name	Age
Robert Tait McKenzie	58
Heikki Ilmari Savolainen	39
Joseph "Josy" Stoffel	38
Ioannis Theofilakis	36
Takashi Ono	33
...	...
Tatyana Vasilyevna Kalmykova	1
Mariya Lvovna Kalmykova	1
Christine Kalmer	1
Joannis "Jannis" Kalmazidis	1
Pierre-Georges LeClercq	1

Name: count, Length: 134732, dtype: int64      Name: count, Length: 74, dtype: int64

Height :

Team :

```

Height
180.0    12492
170.0    11976
178.0    10708
175.0    10320
183.0     8284
...
226.0      3
131.0      2
130.0      2
219.0      2
128.0      1
Name: count, Length: 95, dtype: int64
...

Team
United States    17847
France           11988
Great Britain    11404
Italy             10260
Germany           9326
...
Briar             1
Hannover          1
Nan-2             1
Brentina          1
Digby             1
Name: count, Length: 1184, dtype: int64
...

Games :

```

```

Games
2000 Summer    13821
1996 Summer    13780
2016 Summer    13688
2008 Summer    13602
2004 Summer    13443
1992 Summer    12977
2012 Summer    12920
1988 Summer    12037
1972 Summer    10304
1984 Summer     9454
1976 Summer     8641
1968 Summer     8588
1952 Summer     8270
...
1956 Winter     1307
1904 Summer     1301
1960 Winter     1116
1952 Winter     1088
1948 Winter     1075
1936 Winter      895
1928 Winter      582
1924 Winter      460
1896 Summer      380
1932 Winter      352
... Name: count, dtype: int64

```

On peut alors déduire le type des variables qui compose le jeu de données :

- Variables qualitatives nominales :

ID, Name, Sex, Team, NOC, Games, Season, City, Sport, Event : Ces variables sont toutes qualitatives nominales car elles représentent des catégories distinctes sans ordre spécifique. Par exemple, "Sex" pourrait avoir des valeurs comme "Male" ou "Female", "Team" pourrait être "USA", "NOC" pourrait être "USA" également, "Games" pourraient être "2016 Summer", et ainsi de suite. Aucun ordre spécifique ne peut être établi entre ces catégories, elles sont simplement des labels pour identifier différentes entités ou événements.

- Variable qualitative ordinale :

Medal : Cette variable est qualitative ordinale car elle a un ordre spécifique mais les distances entre les catégories ne sont pas nécessairement uniformes. Par exemple, une médaille d'or est généralement considérée comme ayant une "valeur" supérieure à une médaille d'argent, qui à son tour est considérée comme ayant une "valeur" supérieure à une médaille de bronze. Cependant, la différence entre une médaille d'or et une médaille d'argent peut ne pas être la même que celle entre une médaille d'argent et une médaille de bronze.

- Variable quantitative discrète :

Year : Cette variable est quantitative discrète car elle représente des valeurs numériques qui sont des nombres entiers et qui sont comptées individuellement dans une séquence ordonnée. Chaque année est une unité distincte et ne peut être subdivisée en parties plus petites. Par exemple, 2000, 2001, 2002, etc., sont toutes des années distinctes sans valeurs intermédiaires.

- Variables qualitatives continues :

Height, Weight : Ces variables sont qualitatives continues car elles représentent des mesures numériques qui peuvent avoir une gamme infinie de valeurs. Par exemple, "Height" pourrait être mesurée en centimètres et "Weight" en kilogrammes. Ces variables peuvent prendre n'importe quelle valeur dans une plage donnée et peuvent être mesurées avec précision à n'importe quel niveau de décimales, les rendant continues.

Etudions une variable de chaque type :

- Sport (variable quantitative nominale) :

Effectifs associés :

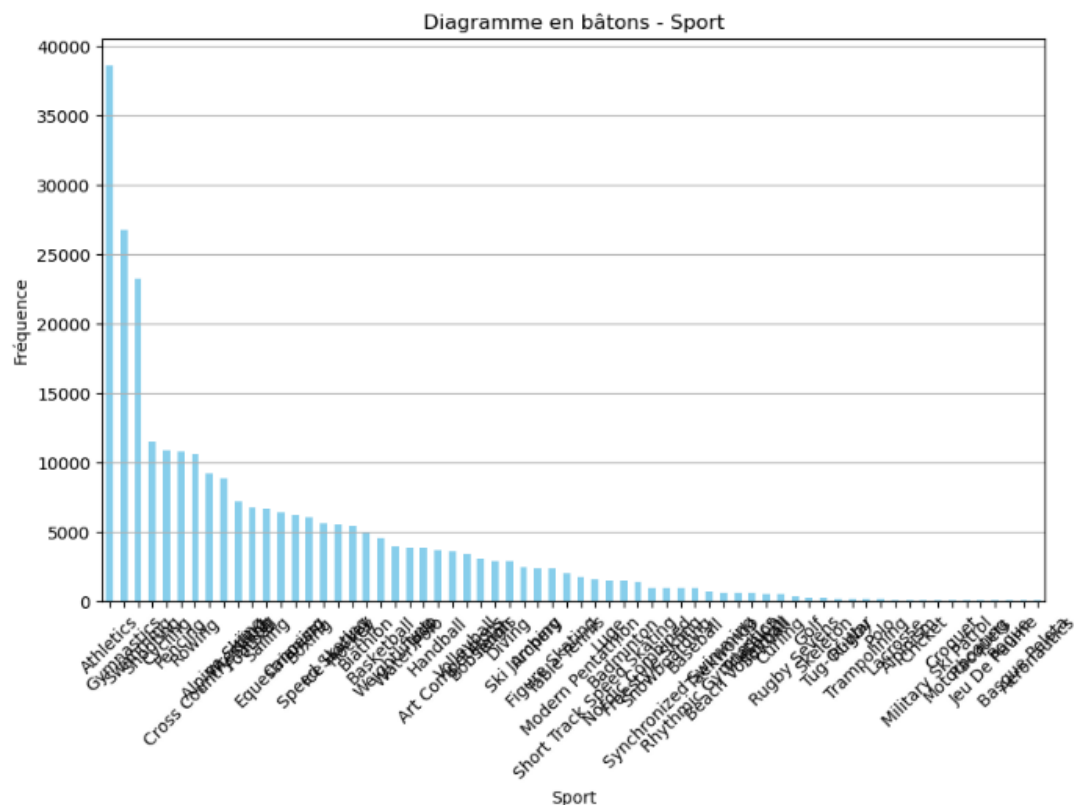
Fréquences associées:

```
Sport
Athletics      38624
Gymnastics     26707
Swimming       23195
Shooting       11448
Cycling        10859
...
Racquets       12
Jeu De Paume   11
Roque          4
Basque Pelota  2
Aeronautics    1
Name: count, Length: 66, dtype: int64
```

Mode de la série :

```
Mode de la série: 0    Athletics
Name: Sport, dtype: object
```

Diagramme en bâton :



- Medal (variable qualitative ordinale) :

Effectifs associés :

```
Fréquences associées:
Medal
Gold      13372
Bronze    13295
Silver    13116
Name: count, dtype: int64
```

Mode de la série :

```
Mode de la série: 0    Gold
Name: Medal, dtype: object
```

- Year (variable quantitative discrète) :

Effectifs associés :

```
Fréquences associées à l'année:
Year
1896      380
1900     1936
1904     1301
1906     1733
1908     3101
1912     4040
1920     4292
1924     5693
1928     5574
1932     3321
1936     7401
1948     7480
1952     9358
1992     16413
1994     3160
1996     13780
1998     3605
2000     13821
2002     4109
2004     13443
2006     4382
2008     13602
2010     4402
2012     12920
2014     4891
2016     13688
... Name: count, dtype: int64
```

Effectifs cumulés :

```
Effectifs cumulés associés à l'année:
Year
1896      380
1900     2316
1904     3617
1906     5350
1908     8451
1912    12491
1920    16783
1924    22476
1928    28050
1932    31371
1936    38772
1948    46252
1952    55610
1992    165313
1994    168473
1996    182253
1998    185858
2000    199679
2002    203788
2004    217231
2006    221613
2008    235215
2010    239617
2012    252537
2014    257428
2016    271116
... Name: count, dtype: int64
```

Fréquences cumulées :

```
Year
1896    0.001402
1900    0.008542
1904    0.013341
1906    0.019733
1908    0.031171
1912    0.046073
1920    0.061903
1924    0.082902
1928    0.103461
1932    0.115711
1936    0.143009
1948    0.170599
1952    0.205115
1992    0.609750
1994    0.621406
1996    0.672233
1998    0.685529
2000    0.736508
2002    0.751663
2004    0.801247
2006    0.817410
2008    0.867581
2010    0.883817
2012    0.931472
2014    0.949512
2016    1.000000
... Name: count, dtype: float64
```

Diagramme en bâtons des effectifs :

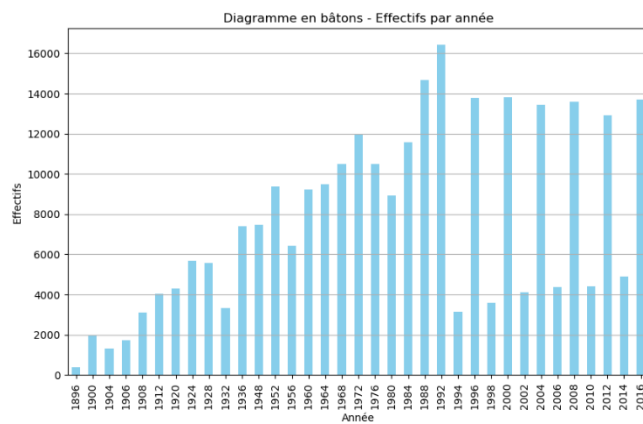
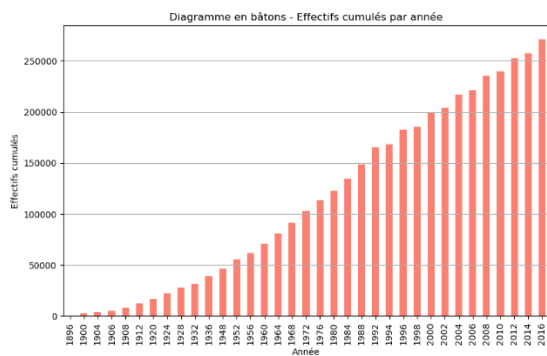
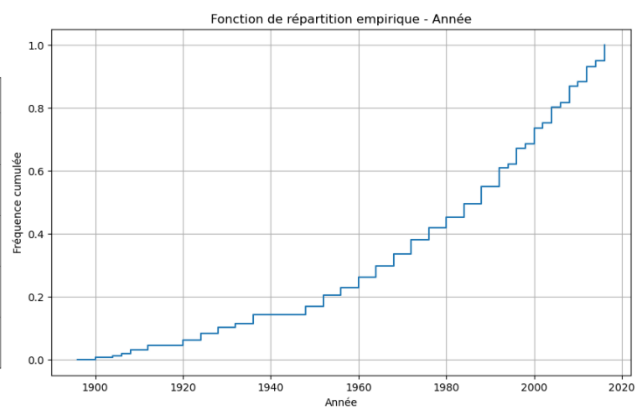


Diagramme circulaire des effectifs cumulés :



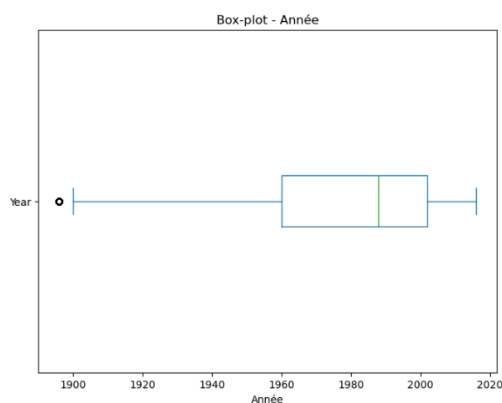
Fonction de répartition :



Statistiques d'ordre principaux, de tendance centrale et de dispersion :

```
Minimum de l'année: 1896
Maximum de l'année: 2016
Médiane de l'année: 1988.0
Mode de l'année: 0 1992
Name: Year, dtype: int64
Écart-type de l'année: 29.877631985612652
Variance de l'année: 892.6728930677043
```

Box-plot :



Nombre d'outliers pour la variable :

Nombre d'outliers pour l'année: 380

- Weight (variable quantitative continue) :

Nombre de classes :

Nombre de classes (règle de Sturges) : 19

Effectifs :

```

Weight
(24.811, 34.947]      250
(34.947, 44.895]     2731
(44.895, 54.842]    19471
(54.842, 64.789]    51330
(64.789, 74.737]    59146
(74.737, 84.684]    43153
(84.684, 94.632]    20561
(94.632, 104.579]   7474
(104.579, 114.526]  2086
(114.526, 124.474]  1061
(124.474, 134.421]  635
(134.421, 144.368]  197
(144.368, 154.316]   84
Name: count, dtype: int64

```

Effectifs cumulés :

```

Effectifs cumulés par classe :
Weight
(24.811, 34.947]      250
(34.947, 44.895]     2981
(44.895, 54.842]    22452
(54.842, 64.789]    73782
(64.789, 74.737]   132928
(74.737, 84.684]   176081
(84.684, 94.632]   196642
(94.632, 104.579]  204116
(104.579, 114.526] 206202
(114.526, 124.474] 207263
(124.474, 134.421] 207898
(134.421, 144.368] 208095
(144.368, 154.316] 208179
(154.316, 164.263] 208218
(164.263, 174.211] 208230
(174.211, 184.158] 208237
(184.158, 194.105] 208238
(194.105, 204.053] 208239
(204.053, 214.0]   208241
Name: count, dtype: int64

```

Fréquences cumulées :

```

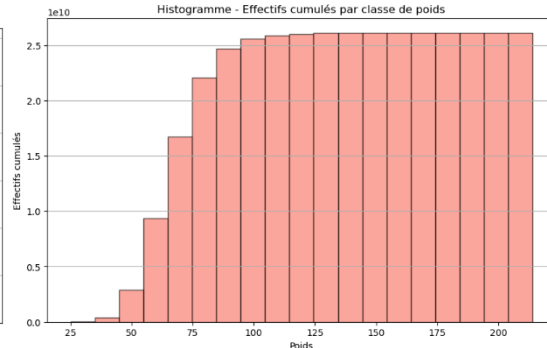
Fréquences cumulées par classe :
Weight
(24.811, 34.947]      0.000922
(34.947, 44.895]     0.010995
(44.895, 54.842]     0.082813
(54.842, 64.789]     0.272142
(64.789, 74.737]     0.490299
(74.737, 84.684]     0.649467
(84.684, 94.632]     0.725306
(94.632, 104.579]    0.752873
(104.579, 114.526]   0.760567
(114.526, 124.474]   0.764481
(124.474, 134.421]   0.766823
(134.421, 144.368]   0.767550
(144.368, 154.316]   0.767860
(154.316, 164.263]   0.768003
(164.263, 174.211]   0.768048
(174.211, 184.158]   0.768073
(184.158, 194.105]   0.768077
(194.105, 204.053]   0.768081
(204.053, 214.0]     0.768088
Name: count, dtype: float64

```

Histogramme des effectifs :

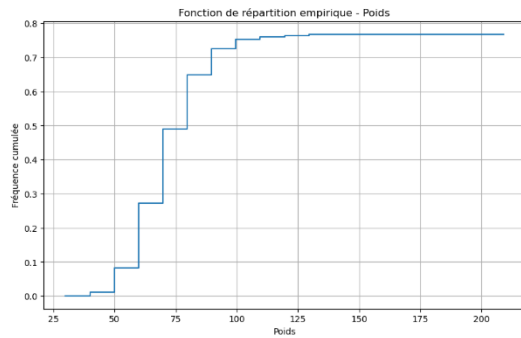


Histogramme des effectifs cumulés :



Fonction de répartition : Statistiques d'ordre principaux, de tendance centrale et de dispersion :





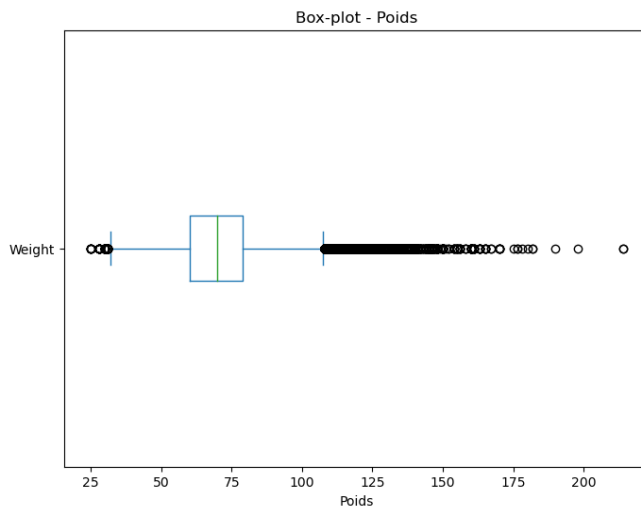
Statistiques d'ordre principaux :  
 Minimum : 25.0  
 Maximum : 214.0

Statistiques de tendance centrale :  
 Moyenne : 70.70239290053351  
 Mode : 70.0  
 Médiane : 70.0

Statistiques de dispersion :  
 Écart-type : 14.348019999019845  
 Variance : 205.86567789227345

Box-plot :

Nombre d'outliers pour la variable :



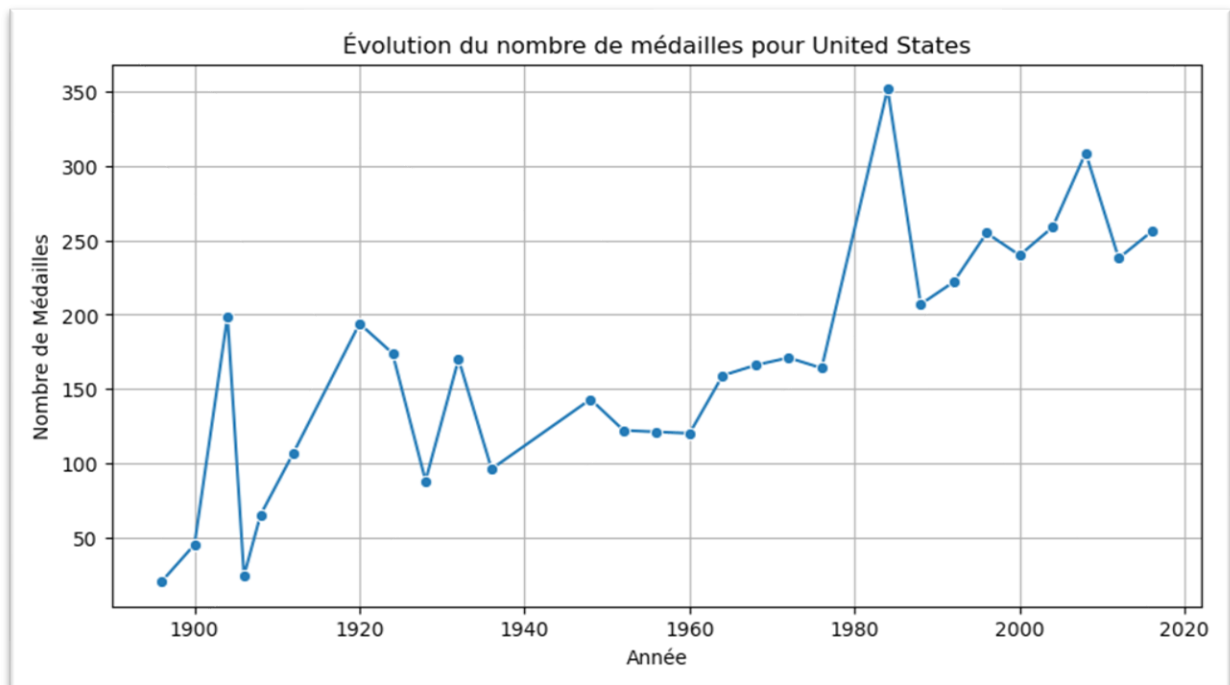
Nombre d'outliers pour le poids: 3237

## 2. Etude de l'évolution du tableau des médailles à travers le temps

Pour cette problématique nous allons explorer comment les performances de certaines équipe nationales ont changé, progressé, ou même régressé, dans le temps. Cette analyse ne se limite pas simplement à compter les médailles, elle cherche à déchiffrer les histoires cachées derrière les chiffres.

### Les Etats-Unis

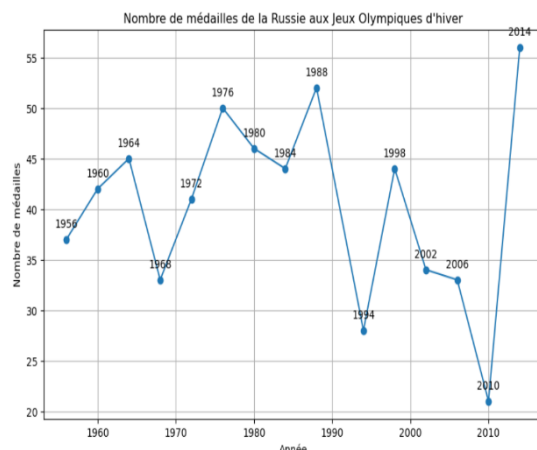
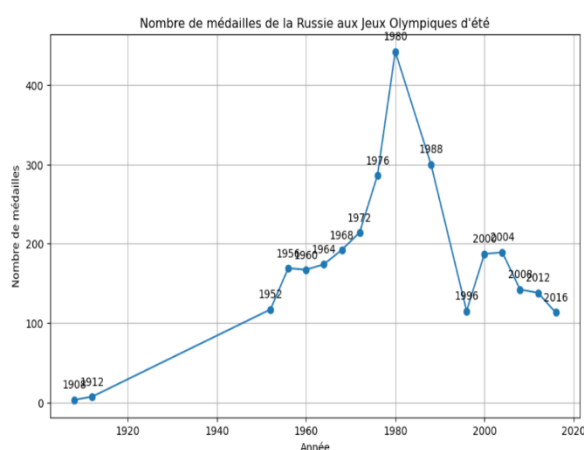
Ici on peut voir par exemple l'évolution des États-Unis pour les jeux olympiques d'été depuis 1896



On constate sur ce graphique que les Etats-Unis ont participé à pratiquement tous les Jeux Olympiques modernes depuis leur création. Les seules olympiades auquel les Etats-Unis n'ont pas participé sont celles de 1980, en effet en pleine guerre froide ces jeux avaient lieu en URSS, les Etats-Unis ont donc décidés de boycotter cette édition.

Parlons maintenant des pays qui pourraient avoir les évolutions les plus flagrantes au cours du temps. Le premier pays auquel on pourrait penser est probablement la Chine. Mais en réalité, le pays avec la croissance la plus forte est la Russie/Union Soviétique, leurs progrès au fil de temps est incroyable autant pendant les jeux olympiques d'été que d'hiver, ils se sont démarqués.

## La Russie

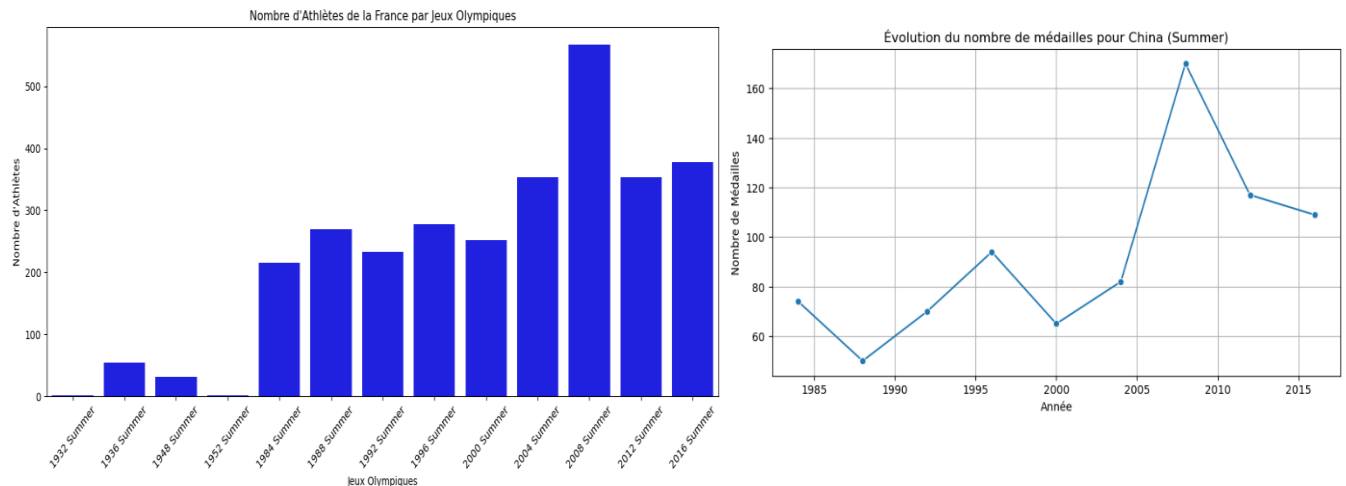


La très forte croissance entre 1952 et 1988 est dû au fait que l'URSS, vainqueurs de la seconde Guerre Mondiale aux côtés des alliés connaît une expansion phénoménale dans tous les domaines, y-compris le sport. Durant la guerre froide, le sport est un des domaines

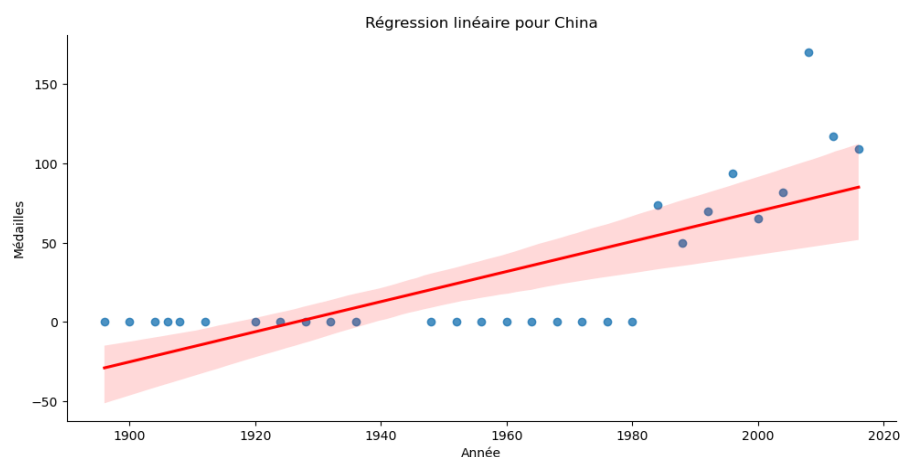
principaux où les Etats-Unis et l'URSS s'opposent, ce qui explique cette croissance. On peut voir un pic en 1980, il est dû au fait qu'en 1980, les Jeux sont organisés en Russie et les USA décident alors de boycotter l'événement, la Russie en profite alors et fait un carton plein. La chute de médailles après 1992 est dû à la dissolution de l'Union Soviétique. Après ces événements, la Russie a poursuivi sa croissance dans le but de maintenir une présence influente sur la scène sportive internationale à long terme. Malgré les défis économiques et politiques internes, la Russie a poursuivi ses investissements dans le développement des infrastructures sportives ainsi que dans la formation des athlètes de haut niveau. Une partie de la motivation derrière ce fort accent sur les sports, en particulier les sports d'hiver où la Russie excelle traditionnellement, est le désir de maintenir un statut de superpuissance dans le domaine sportif, rivalisant avec des nations telles que les États-Unis. Renforcer le patriotisme et l'identité nationale russe est aussi considéré comme un moyen d'unifier le pays autour de symboles de fierté nationale et de réussite internationale.

Néanmoins, il est important de noter que de nombreux indices ont démontré l'usage de produits dopants au fil des années pour les athlètes russes.

La Chine est un des pays dont la croissance des performances est la plus élevée au niveau des performances olympiques. Voilà un diagramme en bâtons montrant le nombre de participants des athlètes chinois pendant les jeux olympiques et à côté l'évolution du pays aux jeux olympiques.



Ce graphique montre clairement l'évolution de la Chine dans les différents sports au niveau mondial.



Cette analyse nous montre à quel point la géopolitique et les Jeux Olympiques sont fortement liés et donc à quel point il est parfois difficile d'interpréter des résultats. Dans la suite de notre étude nous différencierons parfois les résultats des Jeux Olympiques avant 1992 et ceux après cette date, pour les raisons suivantes :

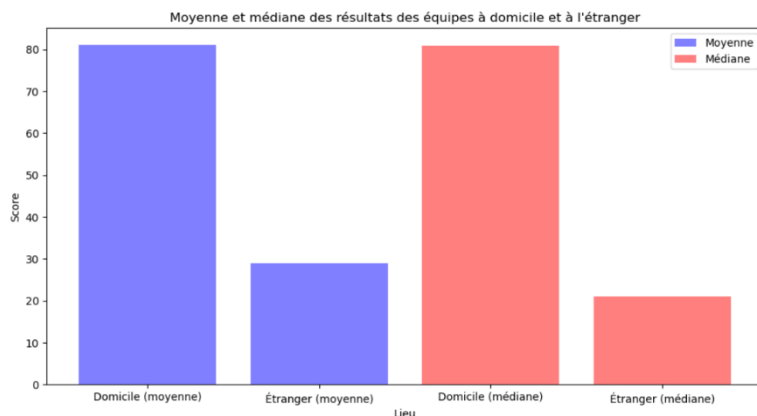
- Guerres Mondiales et guerre froide : les différents pays qui participent aux Jeux, peuvent avoir un statut économique avant pendant et après ces guerres, totalement différent. De plus, l'Allemagne était, durant la guerre froide séparée en deux (Est et Ouest) et certains jeux étaient boycottés soit pas les USA et leurs alliés soit par les Russes et les leurs.
- Mondialisation : prenons un exemple parlant : en 1900 les JO se passaient en France : 679 athlètes représentaient alors la France et 114 les USA. En 1904 les JO se passaient aux USA et 1 seul athlète Français avait alors fait le déplacement et 823 athlètes concouraient pour les USA. Les athlètes français n'ont pas participé à ces jeux non pas parce qu'ils n'étaient pas qualifiés mais simplement parce qu'en 1900 l'unique moyen de transport transatlantique était le bateau dont le prix des billets était très élevé et le voyage durait quelques semaines. Tandis que de nos jours tout athlète sélectionné pour les JO ne rencontre aucun problème pour se rendre sur lieu de la compétition. Cet argument peut aussi simplement se prouver par le fait qu'en 1896, 14 nations différentes étaient représentées tandis qu'en 2016, 204 équipes (pays et délégations) participaient.

### 3. Etude des performances des équipes en fonction de si elles sont à domicile ou non

Avec les Jeux Olympiques de Paris 2024 on peut se demander si nous avons le droit d'espérer obtenir plus de médailles cette année puisque nos athlètes évolueront à domicile. Nous allons donc dans cette partie, étudier les différences de performances des athlètes olympiques en fonction de s'ils sont à domicile ou non.

#### Analyse brute des données

Voici un graphique montrant que l'équipe représentant le pays hôte des Jeux Olympiques performe bien mieux que lorsque celui-ci évolue à l'étranger.

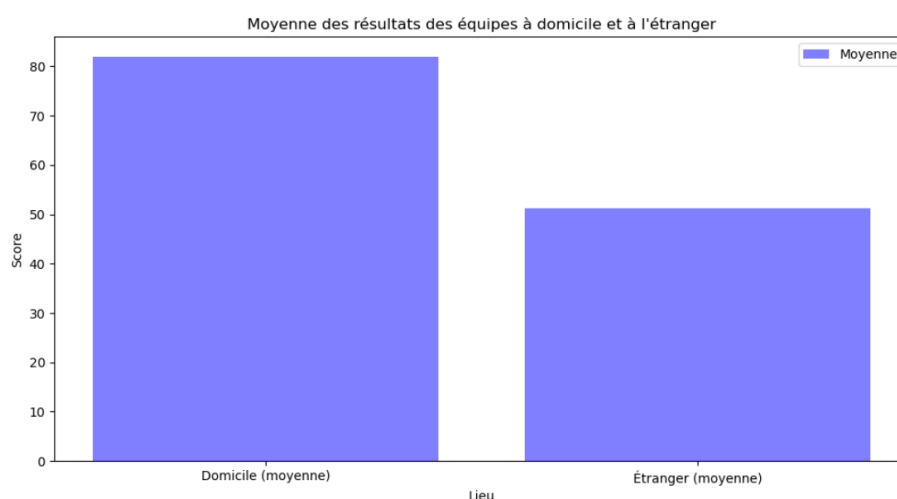


Dans ce graphique, on peut retrouver le nombre de médailles moyen et médian obtenu par les équipes lorsque celles-ci évoluaient à domicile, ainsi que le nombre moyen et médian de ces équipes lorsque celles-ci évoluaient à l'étranger, le tout depuis 1896. On constate que les médianes et les moyennes sont pratiquement confondues c'est

pourquoi nous n'afficheront que les moyennes pour les graphiques suivants. On remarque dans ce graphique que la différence est frappante entre les deux cas. En effet, il y a en moyenne environ 80 athlètes médaillés pour les équipes à domicile contre 30 lorsque ces mêmes équipes participent à des jeux à l'étranger, cela fait une différence de 50 médaillés, ce qui est énorme !

### Nuancement

Néanmoins, comme expliqué dans l'étude de l'évolution du tableau des médailles à travers le temps, on sait que la situation géopolitique du monde a énormément évolué depuis 1896, c'est pourquoi nous avons décidé d'aussi regarder les résultats après 1992.



Ce graphique est le même que le précédent mais cette fois il ne prend en compte que les résultats des équipes depuis 1992.

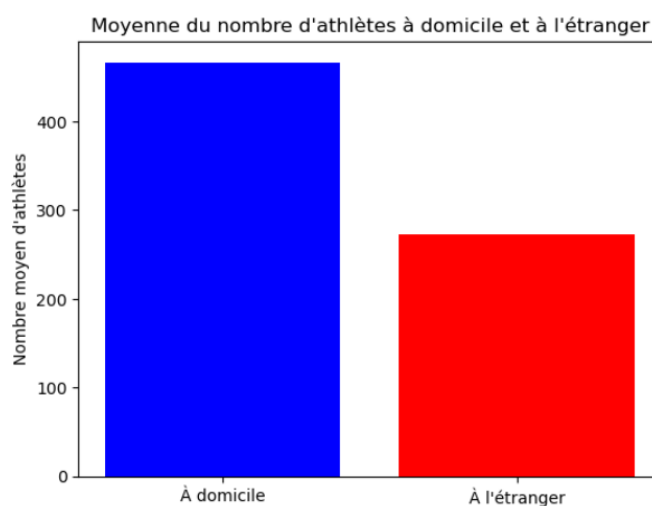
Ce diagramme nous montre deux choses :

- La première, que les équipes qui évoluent à domicile obtiennent bel et bien toujours plus d'athlètes médaillés que lorsque celles-ci évoluent à l'étranger.
- La seconde, que la différence est bien moins évidente mais si elle reste conséquente, en effet, ici nous avons une différence de 30 athlètes médaillés en plus en moyenne lorsqu'une équipe évolue à domicile que lorsque celle-ci évolue à l'étranger contre 50 pour le diagramme précédent

Cette différence s'explique notamment par l'évolution du sport dans la société depuis 1896. De fait, en 1896 le sport n'avait pas du tout la même place dans la société qu'aujourd'hui, à cette époque être sportif n'était pas un métier, c'était un passe-temps, un hobby. Donc lorsque des jeux étaient organisés à domicile, c'était simple de s'y rendre, simple d'y participer et donc d'y performer. Mais lorsqu'il fallait aller à l'autre bout du monde pour pouvoir participer à des olympiades c'était bien plus difficile. En outre c'était bien plus coûteux qu'aujourd'hui le prix des billets de bateau pour traverser l'océan étaient accessibles uniquement pour une classe aisée mais on peut aussi rajouter le fait qu'un aller-retour à l'autre bout du monde pouvait facilement durer un mois entier. Or, à cette époque il est inconcevable pour un travailleur de quitter son travail un mois durant et en brûlant toutes ses économies pour des jeux olympiques. Pour appuyer cet argument voici un exemple parlant. En 1900, les jeux sont organisés à Paris, en France. Le nombre d'athlètes français ayant participé en 1900 s'élève à 679 et la France obtient cette année-là 75 athlètes médaillés. En 1904, 4 ans plus tard, les jeux sont organisés à Saint-Louis, aux Etats-Unis. Un seul athlète français fera le déplacement aux Etats-Unis afin de défendre les couleurs de la France (fièrement puisque celui-ci obtiendra une médaille).

#### Explications des résultats de l'analyse

Nous avons vu que les équipes olympiques gagnaient plus de médailles lorsqu'ils sont à domicile que lorsqu'ils ne le sont pas. Il y a plusieurs raisons à cela. La première n'est pas la plus évidente mais celle qui a le plus fort impact sur les résultats. Tout simplement lorsqu'un pays organise une olympiade, toutes les équipes nationales représentant ce pays dans les sports collectifs sont qualifiées d'office et n'ont pas à passer de matchs de barrages, etc. De plus, pour certains sports comme le tir à l'arc, l'équipe du pays hôte ont le droit d'inscrire le nombre maximum d'athlète dans leur équipe alors que pour les autres équipes cela se joue sur le classement mondial de leurs athlètes. Enfin, pour d'autres sports individuels comme l'athlétisme l'équipe du pays organisateurs peut avoir droit à un quota minimum d'athlètes qualifiés même si ceux-ci n'ont pas réalisé les minimas olympiques.



Ce diagramme prouve le point précédent. On voit ici que la moyenne des athlètes participant à des olympiades dans leur propre pays est d'approximativement 465 alors que seulement environ 270 athlètes en moyenne participent aux Jeux lorsqu'ils sont à l'étranger.

Une autre raison plus attendue serait la motivation des athlètes à performer à domicile, en réalité les athlètes ont souvent plus à cœur de performer lors de jeux à domicile qu'à l'étranger et sont plus susceptibles de faire plus de sacrifices pour performer lors de jeux à

domicile que lors de jeux à l'étranger. Il ne faut pas oublier non plus le facteur du public ! En effet des athlètes évoluant à domicile sont bien plus encouragés ce qui peut influencer sur leur mental lors d'une course ou d'un match très serré. Autre faveur pour le pays hôte : la possibilité de choisir quatre des 32 disciplines au programme, et donc maximiser ses chances de médailles. Le reste des disciplines sont choisies par le comité olympique international. Enfin, pour des jeux à domicile le pays hôte va potentiellement investir plus dans ses athlètes de haut niveau. Par exemple elle peut construire de nouvelles installations qui vont permettre aux sportifs de mieux pouvoir s'entraîner.

#### 4. Etude des performances des athlètes en fonction leur âge et de leur sexe

Pour étudier cette problématique, nous nous sommes intéressés en premier lieu à la moyenne d'âge des médaillés, masculins et féminins pour avoir une première idée sur la réponse à notre problématique. Cependant, une moyenne arithmétique à elle seule ne permet pas de conclure notre raisonnement mais intéressons-nous déjà à ces valeurs.

##### Âge moyen des athlètes médaillés :

##### Âge moyen des athlètes masculins médaillés :

###### Medal

**Bronze 26.355729**

**Gold 26.504641**

**Silver 26.633194**

On remarque que pour les hommes, les médaillés depuis le début des jeux olympiques en 1896 ont en moyenne 26ans et demi. On remarque aussi qu'il n'y a pas de différence significative d'âge entre les différents types de médailles.

##### Âge moyen des athlètes féminins médaillées :

###### Medal

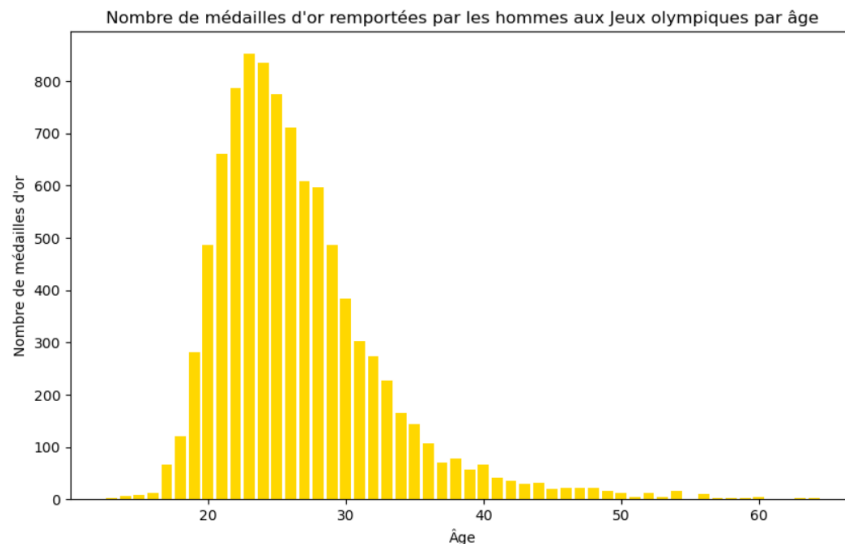
**Bronze 24.708743**

**Gold 24.372596**

**Silver 24.444891**

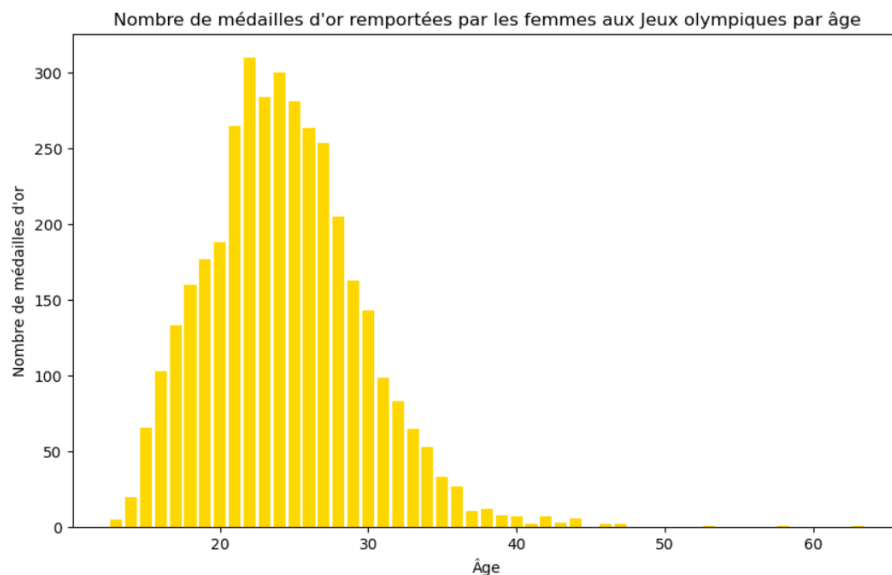
On remarque que pour les femmes, les médaillées depuis le début des jeux olympiques en 1924 ont en moyenne 24ans et demi. Comme pour les hommes, on remarque qu'il n'y a pas de différence significative d'âge entre les différents types de médailles.

Comme évoqué précédemment, une moyenne arithmétique permet de répondre à notre question uniquement si l'âge des médaillés suit une distribution normale. Or pour le moment, nous avons aucune information à ce sujet. Cependant, un graphique à barre liant l'âge au nombre de médailles peut nous aider, essayons de générer ce graphe.



On remarque que notre variable ne suit pas une distribution normale à vue d'œil donc la moyenne arithmétique ne peut pas répondre parfaitement à notre problématique. Néanmoins si on considère la variable âge comme quantitatif discrète comme dans le jeu de donnée, les hommes performant le mieux à **23 ans**.

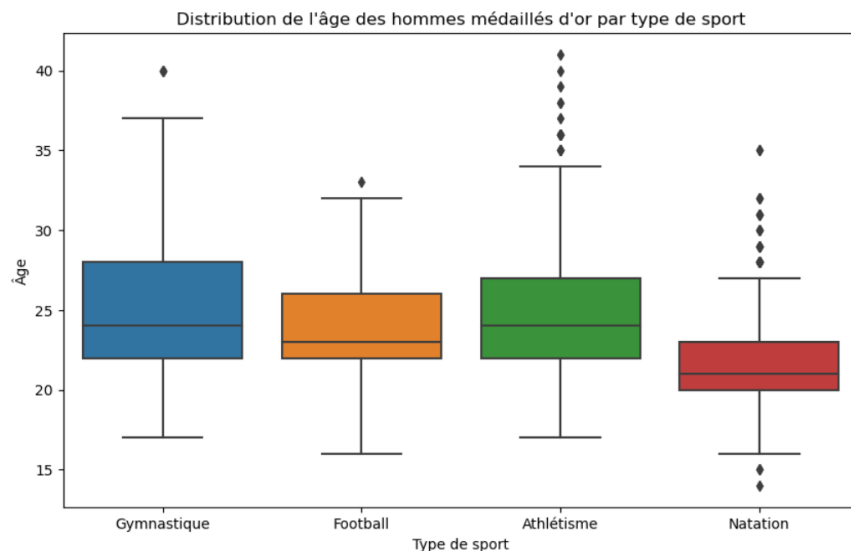
Vérifions si on arrive à la même déduction pour les femmes :



Comme pour les hommes, notre variable ne suit pas une distribution normale à vue d'œil, on arrive donc à la même conclusion. Cette fois ci, en considérant la variable âge comme quantitatif discrète comme dans le jeu de donnée, les femmes performant le mieux à **22 ans**.

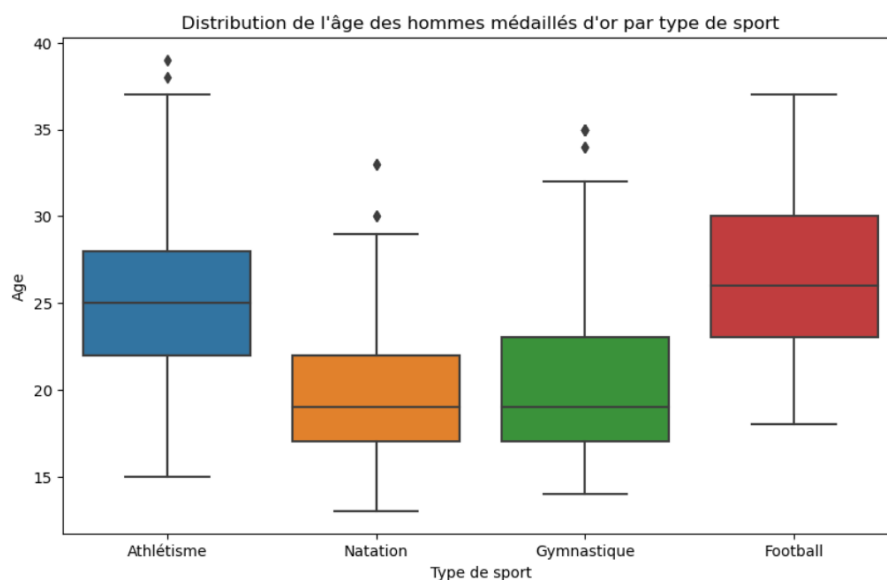
Intéressons-nous maintenant à notre problématique dans des sports en particulier. Pour notre étude, nous avons choisi : le football, la gymnastique, la natation et l'athlétisme.





On remarque qu'en fonction des sports, l'âge des hommes qui performent le mieux est différent. Dans notre cas, la natation est le sport qui se distingue le plus parmi les 4, en effet, parmi les quatre, c'est le sport où les médaillés sont les plus jeunes avec une médiane proche de 20 ans et les 3 quarts des médaillés ayant moins de 23 ans.

Intéressons-nous maintenant à cette problématique mais du point de vue des femmes :



On remarque que pour les femmes, la différence d'âge des médaillés en fonction des 4 sports choisis est plus grande que pour les hommes. Par exemple, pour l'athlétisme, la médiane est à 25 ans alors que pour la natation et la gymnastique, la médiane est située légèrement en dessous de 20 ans. On peut donc en déduire que le sport, que ce soit pour les hommes ou pour les femmes, influence sur l'âge moyen des sportifs qui performent le mieux aux JO.

Pour exploiter la régression linéaire multiple, intéressons-nous à l'évolution du nombre de médailles d'or remporté par les hommes par à l'âge et à la taille de ces derniers.

La fonction prédéfinie de python summary nous donne :

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Medal    R-squared:                0.002
Model:                  OLS      Adj. R-squared:           0.002
Method:                 Least Squares    F-statistic:             19.81
Date:                  Tue, 30 Apr 2024    Prob (F-statistic):       2.53e-09
Time:                  15:48:27    Log-Likelihood:          -36835.
No. Observations:      20571    AIC:                     7.368e+04
Df Residuals:          20568    BIC:                     7.370e+04
Df Model:               2
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	0.0281	0.191	0.147	0.883	-0.346	0.402
Age	0.0112	0.002	5.613	0.000	0.007	0.015
Height	0.0028	0.001	2.803	0.005	0.001	0.005

D'après ces résultats, il semble que l'âge et la taille ont un impact significatif sur le nombre de médaille d'or remporté par les hommes. Cependant, il semble que cette relation n'est pas linéaire ce qui est cohérent avec notre précédente étude.

#### Réponse à la problématique

En réponse à notre problématique, nous pouvons conclure que l'âge optimal pour performer aux Jeux Olympiques varie en fonction du genre et du sport pratiqué. Globalement, pour les hommes, l'âge de performance optimale semble être autour de 23 ans, tandis que pour les femmes, il semble être légèrement inférieur, autour de 22 ans. Cependant, cet âge idéal est trop général et ne prend pas en compte le fait que le sport pratiqué influe sur l'âge comme a pu le voir avec la natation ou l'athlétisme. Finalement, notre analyse sur cette problématique nous a permis d'avoir une idée générale sur la réponse mais nous a surtout montré la complexité du lien entre l'âge et la performance aux jeux olympiques.

#### IV. Conclusion

Nous arrivons à la fin de notre analyse des possibles facteurs qui pouvaient jouer sur les performances des athlètes qui vont nous représenter cet été lors des Jeux Olympiques de Paris 2024. Dans notre étude nous avons vu qu'il y avait énormément de facteurs qui pouvaient jouer. Nous avons aussi vu que les Jeux Olympiques étaient très reliés à la géopolitique et aux autres événements majeurs qui pouvaient se produire. De plus, nous avons vu que beaucoup de facteurs semblaient jouer en faveur de l'amélioration des performances des équipes à domicile, comme les qualifications d'office, la motivation des athlètes ou encore l'investissement des pays hôtes. Enfin nous avons aussi parlé de l'âge, en effet l'âge est un facteur important pour un sportif de haut-niveau et nous sommes arrivés à la conclusion que les athlètes olympiques performaient le mieux à 23 ans pour les hommes et 22 ans pour les femmes.