TP L3 : Pandas

Sur Moodle, est déposé un répertoire contenant un fichier intitulé « 120-years-of-olympic-history-athletes-and-results.csv ». Il contient des informations sur les résultats de l'ensemble des athlètes ayant participé aux Jeux olympiques de 1896 à 2016.

1) Leire le fichier CSV
2) Décrire la base de données : nombre de lignes, de colonnes, leur signification
3) Gestion de l'en-tête : supprimer l'entête en anglais et remplacer là par une entête en français
4) Définir la première colonne du fichier comme étant l'index du tableau.
5) Sélectionner uniquement les colonnes contenant les noms des athlètes, leur poids et leur taille en utilisant les positions des colonnes, puis en utilisant leur nom.
6) Explorer les types de données du dataframe
7) Si des colonnes contiennent des dates, utiliser la méthode « parser » pour formater correctement ces dates.
8) Lire le fichier Excel
9) Lire la base de données
10) Lire le jeu de données JSON
11) Sélectionner la colonne qui correspond au poids des athlètes du tableau « 120 years … », et afficher la.
12) Demander à Pandas de sélectionner les colonnes 1 et 5, respectivement Name et Weight, ave l'option usecols.
    Spécifier aussi à Pandas qu'on souhaite que l'objet créé soit de classe Series, grâce à l'option squeeze, et définir que la première colonne de cet objet (la colonne Name ici) correspondra aux index, ave l'option index_col.
13) Obtenir les poids des athlètes « Antti Sami Aalto » et « Andrzej ya ».
14) Créer une série, qui contient quatre valeurs et dont les noms d'index sont respectivement 1, 3, 2 et 0.
15) Sélectionner uniquement les athlètes avec un poids strictement supérieur à 90 kg.
16) Même question que ci-dessus avec des poids strictement supérieurs à 90 et strictement inférieures à 100.
17) Comment afficher le tableau suivant :

```
1000    84.0
1001    84.0
1002    84.0
1003    73.0
1004    54.0
        ...
271111  89.0
271112  59.0
271113  59.0
271114  96.0
271115  96.0
Name: Weight, Length: 270116, dtype: float64
```

18) Comment afficher le tableau suivant :

```
0     80.0
1     60.0
2      NaN
3      NaN
4     82.0
5     82.0
6     82.0
7     82.0
8     82.0
9     82.0
10    75.0
11    75.0
12    75.0
13    75.0
14    75.0
15    75.0
16    75.0
17    75.0
18    72.0
19    72.0
Name: Weight, dtype: float64
```

19) Comment déterminer la taille de notre série contenant le poids des athlètes.
20) Obtenir différentes statistiques de la série donnant les poids des athlètes.
21) Indexer le dataframe en choisissant la colonne des noms comme index, et afficher les données de Usain Bolt.
22) Comparer les résultats d'Usain Bolt ave ceux de son adversaire, Justin Alexander Gatlin
23) Sélectionner l'ensemble des femmes ayant participé aux JO et ayant gagné une médaille d'or.
24) Donner le nombre de valeurs manquantes par variable
25) Supprimer les athlètes dont la taille n'est pas répertoriée dans le jeu de données.
26) Remplacer les valeurs manquantes des colonnes Age, Height et Weight par la valeur moyenne de chaque colonne, respectivement.
27) Détecter les lignes dupliquées. Combien de fois sont-elles dupliquées ?
28) Supprimer les lignes dupliquées. Combien reste-t-il de lignes ?
29) Afficher le shape (forme) du dataframe obtenu après suppression des lignes dupliquées.

30) Exécuter :
    `donnees.groupby("Sport").mean().loc[:,['Age','Height','Weight']]`
    Expliquer le dataframe obtenu.
31) Comment obtenir la moyenne d'âge, de taille et de poids par sport et par genre des athlètes.