


Année Universitaire : 2023-2024	Module : Analyse de données pour l'entreprise 1	
Niveau: MRDS1		
Enseignants responsables: Nahla Haddar Ouali		
Note: Ci-joint, avec l'énoncé du projet, vous trouverez le fichier contenant les données à analyser.		

Mini Projet: Classement des universités entre 2017 et 2020

Le classement académique des universités mondiales par l'université Jiao Tong de Shanghai ou classement de Shanghai, est un classement des principales universités mondiales, établi par des chercheurs de l'université Jiao-tong de Shanghai en Chine.

Soit le fichier classement.csv, une base de données qui porte sur le classement de Shanghai des Universités mondiales depuis 2017 jusqu'au 2020:

A	B	C	D	E	F	G	H	I
world_rank	institution	country	national_rank	publications	citations	patents	score	year
1	Harvard Univ	USA	1	1	1	5	100.0	2017
2	Massachusetts	USA	2	12	4	1	91.67	2017
3	Stanford Univ	USA	3	4	2	15	89.5	2017
4	University of	United Kingdom	1	16	11	50	86.17	2017
5	California Institute of	USA	4	37	22	18	85.21	2017
6	Princeton University	USA	5	53	26	101	82.5	2017
7	University of	United Kingdom	2	15	19	26	82.34	2017
8	Yale University	USA	6	14	15	66	79.14	2017
9	Columbia University	USA	7	13	14	5	78.86	2017
10	University of	USA	8	6	3	16	78.55	2017
11	University of	USA	9	34	28	101	73.82	2017

Signification des libellés

- **world_rank** : classement mondial de l'Université
- **institution**: Nom de l'Université
- **country** : Le pays dans lequel se trouve l'Université
- **national_rank**: classement de l'Université au niveau national
- **publications**: Rang en fonction des publications (plus l'Université publie, mieux elle est classée)
- **citations**: Rang en fonction des citations de l'Université
- **patents**: rang en fonction des brevets
- **score**: score total, utilisé pour déterminer le classement
- **year**: année du classement (2017 à 2020)

Les données du fichier sont triées dans l'ordre croissant selon l'année (**year**) et le classement mondial de l'université (**world_rank**).

Travail demandé:

I. Dans le programme principal:

- Créez un DataFrame à partir du contenu du fichier classement.csv
- Affichez combien de citations il y a eu par année et par pays?
- Affichez combien de publications il y a eu par année et par pays.
- Affichez la liste des pays.
- Créez un fichier Excel dont le nom est saisi au clavier. Ce fichier sera utilisé pour sauvegarder les dataframes et les graphiques résultants des fonctions ci-dessous.
- Proposez un menu pour tester les fonctions ci-dessous.

II. Les fonctions: définissez les fonctions suivantes,

1. `save_excel(df, nameF, sheetname)`: permettant de sauvegarder le contenu d'un DataFrame (df) dans une feuille de calcul nommée sheetname, et de l'ajouter au fichier Excel nommé nameF.
2. `save_plot(plt, nameF, sheetname)`: permettant de sauvegarder un graphique dans une feuille de calcul nommée sheetname, et de l'ajouter au fichier Excel nommé nameF.
3. `get_First_Last_University()`: qui demande à l'utilisateur de saisir une année et retourne, pour l'année donnée, un dataframe contenant la première et la dernière université du classement selon world_rank.

Exemple d'exécution:

```
=====> Taper l'année : 2017
First_university Last_university
Harvard University Mines ParisTech
```

4. `get_top10_patents()`: qui retourne, pour une année donnée, un nouveau dataframe contenant les 10 premières universités classées selon 'patents' (rang en fonction des brevets) .

Exemple d'exécution:

```
=====> Taper l'année :
2020
```

	patents	country	institution
1202	1	USA	Massachusetts Institute of Technology
1215	2	USA	Johns Hopkins University
1200	3	USA	Harvard University
1205	4	USA	Columbia University
1259	5	USA	University of Florida
1223	6	South Korea	Seoul National University
1212	7	Japan	University of Tokyo
1343	8	South Korea	Korea Advanced Institute of Science and Techno...
1214	9	USA	University of California, Los Angeles
1201	10	USA	Stanford University

5. `score()` : qui retourne, pour une année donnée (saisie au clavier), deux dataframes, l'un contient es 10 premières universités ayant le maximum de scores, et l'autre contient les 10 premières universités ayant le minimum de scores, groupés par pays.

Exemple d'exécution:

=====> Taper l'année : 2018

Les 10 premiers top scores:

country	institution	score
USA	Harvard University	100.00
United Kingdom	University of Oxford	92.54
Japan	University of Tokyo	76.23
Switzerland	Swiss Federal Institute of Technology in Zurich	64.99
Israel	Hebrew University of Jerusalem	59.98
Canada	University of Toronto	56.11
France	University of Paris-Sud	51.72
South Korea	Seoul National University	51.31
Sweden	Karolinska Institute	47.98
Italy	Sapienza University of Rome	47.75

Les 10 premiers min scores:

country	institution	score
USA	Georgia Institute of Technology	44.26
France	Paris Diderot University - Paris 7	44.36
Finland	University of Helsinki	44.39
Australia	Australian National University	44.50
Canada	University of Alberta	44.50
Japan	Tohoku University	44.52
Russia	Lomonosov Moscow State University	44.93
Singapore	National University of Singapore	45.20
Netherlands	Utrecht University	45.73
Switzerland	Swiss Federal Institute of Technology in Lausanne	45.73

6. `total_ranked_universities()` : qui retourne, pour une année donnée (saisie au clavier), un dataframe contenant le nombre total des universités par pays, le résultat doit être trié du nombre le plus grand au plus petit.

Exemple d'exécution:

=====> Taper l'année : 2020

country	Total_Univ_Par_Pays
USA	229
China	83
Japan	74
United Kingdom	65
Germany	55
France	49
Italy	47
Spain	40
South Korea	36
Canada	33
Australia	27
Taiwan	21
Brazil	18
India	16
Netherlands	13
Austria	12
Sweden	11
Turkey	10
...	...

7. `max_universities()` : qui retourne pour chaque année, la pays ayant le maximum de classements.

Exemple d'exécution:

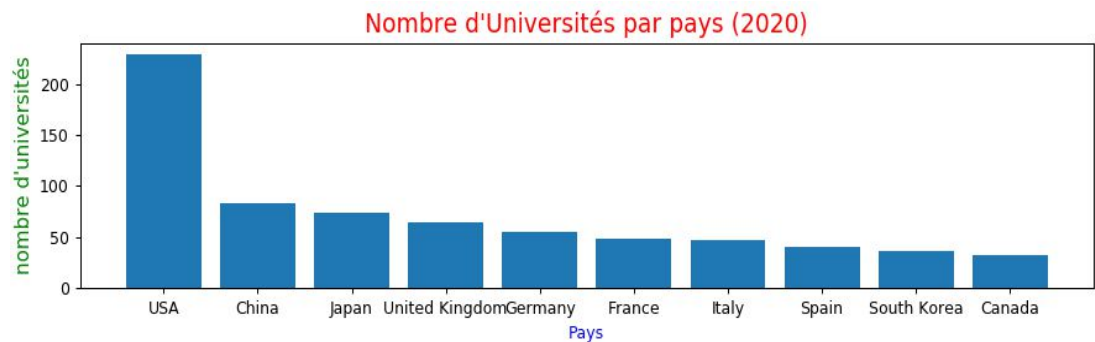
```

Max_Univ_Par_Anne_Pays
year country
2017 USA 58
2018 USA 57
2019 USA 229
2020 USA 229

```

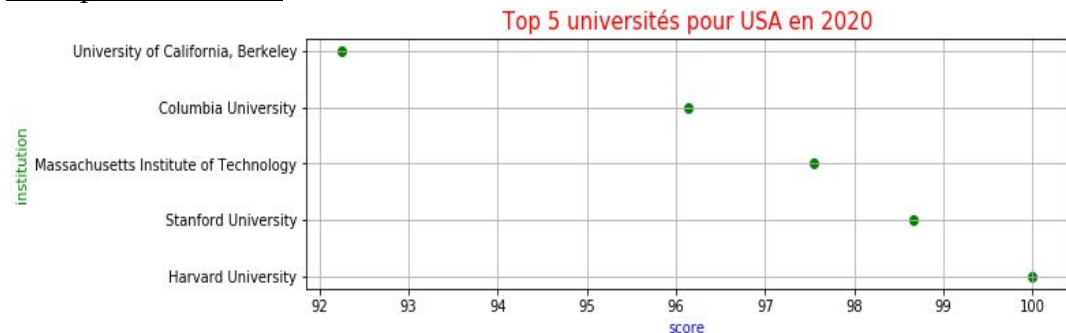
8. `create_bar_graphic()` : qui affiche le nombre total des universités des top 10 pays d'une année donnée (saisie au clavier) , pensez à utiliser le résultat de la fonction `total_ranked_universities`.

Exemple d'exécution:



9. `create_scatter_graphic()` : qui affiche le score des top 5 universités pour un pays et une année donnés (saisis au clavier).

Exemple d'exécution:



NB: que les saisies dans toutes les fonctions doivent être vérifiées, c.à.d. une année doit être entre 2017 et 2020, et le pays doit exister dans la colonne "country" du dataframe.