

## Cas pratiques : Web Scraping avec BeautifulSoup

### Cas pratique 1 :

Web Scraping à partir du Site:

[https://www.basketball-reference.com/leagues/NBA\\_2022\\_per\\_game.html](https://www.basketball-reference.com/leagues/NBA_2022_per_game.html)

Le but est de récupérer les informations contenues dans le tableau comportant les statistiques des joueurs. Insérer les informations extraites dans un Dataframe puis l'exporter dans un fichier.csv. Voici un aperçu du Dataframe souhaité:

	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	Precious Achiuwa	C	22	TOR	73	28	23.6	3.6	8.3	.439	...	.595	2.0	4.5	6.5	1.1	0.5	0.6	1.2	2.1	9.1
1	Steven Adams	C	28	MEM	76	75	26.3	2.8	5.1	.547	...	.543	4.6	5.4	10.0	3.4	0.9	0.8	1.5	2.0	6.9
2	Bam Adebayo	C	24	MIA	56	56	32.6	7.3	13.0	.557	...	.753	2.4	7.6	10.1	3.4	1.4	0.8	2.6	3.1	19.1
3	Santi Aldama	PF	21	MEM	32	0	11.3	1.7	4.1	.402	...	.625	1.0	1.7	2.7	0.7	0.2	0.3	0.5	1.1	4.1
4	LaMarcus Aldridge	C	36	BRK	47	12	22.3	5.4	9.7	.550	...	.873	1.6	3.9	5.5	0.9	0.3	1.0	0.9	1.7	12.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
600	Thaddeus Young	PF	33	TOT	52	1	16.3	2.7	5.2	.518	...	.469	1.5	2.5	4.0	2.0	1.0	0.3	1.0	1.6	6.2
601	Trae Young	PG	23	ATL	76	76	34.9	9.4	20.3	.460	...	.904	0.7	3.1	3.7	9.7	0.9	0.1	4.0	1.7	28.4
602	Omer Yurtseven	C	23	MIA	56	12	12.6	2.3	4.4	.526	...	.623	1.5	3.7	5.3	0.9	0.3	0.4	0.7	1.5	5.3
603	Cody Zeller	C	29	POR	27	0	13.1	1.9	3.3	.567	...	.776	1.9	2.8	4.6	0.8	0.3	0.2	0.7	2.1	5.2
604	Ivica Zubac	C	24	LAC	76	76	24.4	4.1	6.5	.626	...	.727	2.9	5.6	8.5	1.6	0.5	1.0	1.5	2.7	10.3

### Cas pratique 2 :

Web Scraping à partir du Site:

<https://content.codecademy.com/courses/beautifulsoup/cacao/index.html>

Le but est de récupérer les informations contenues dans les deux colonnes « Cocoa percent » et « Rating » du tableau, comportant le pourcentage du chocolat en cacao et son évaluation respectivement. Insérer les informations extraites dans un Dataframe puis l'exporter dans un fichier.csv et un fichier.json. Pour le format json, choisir la valeur index puis records et values pour le paramètre orient et constater la différence. Voici un aperçu du Dataframe souhaité:

	Rating	CocoaPercentage
0	3.75	63.0
1	2.75	70.0
2	3.00	70.0
3	3.50	70.0
4	3.50	70.0
...	...	...
1790	3.75	70.0
1791	3.00	65.0
1792	3.50	65.0
1793	3.25	62.0
1794	3.00	65.0

### Cas pratique 3:

Scraping à partir du site :

<https://www.webscraper.io/test-sites/e-commerce/static/computers/laptops>

Travail demandé :

- Récupérer le nom des articles, la description ainsi que le prix pour tous les laptops dans la page.
- Insérer les données dans un dataframe.
- Récupérer les données des tablets.

Suite :

- Récupérer les informations (nom, description, prix, rating) des laptops de toutes les pages en utilisant BeautifulSoup.
- Insérer les données récupérées dans un dataframe.
- En utilisant le dataframe obtenu, récupérer les informations suivantes (afficher chaque résultat dans un dataframe) :
  1. Les 3 laptops les mieux notés.
  2. Les 3 laptops les moins bien notés.
  3. Les 3 laptops les plus chers.
  4. Les 3 laptops les moins chers.
  5. Refaire le même travail pour les tablets.