

## WEB SCRAPING :SCRAPY :

### *1-web scraping ?*

*Le Web Scraping devient un outil très puissant dans la boîte à outils de la science des données d'aujourd'hui. L'énorme quantité de données structurées et non structurées librement disponibles sur le web peut être utilisée dans divers processus décisionnels. En outre, elles servent souvent de variables cruciales dans de nombreux modèles d'apprentissage automatique et d'intelligence artificielle. Le scraping permet de collecter des informations de nature bien différente. Ces informations sont alors rassemblées dans des bases de données locales ou des tableaux.*

*Nous allons utiliser Anaconda Navigator avec Pycharm pour explorer un site web appelé Amazon : '[https://www.amazon.com/Apple-iPhone-256GB-Silver-T-Mobile/product-reviews/Bo7RV52TRF/ref=cm\\_cr\\_dp\\_d\\_show\\_all\\_btm?/](https://www.amazon.com/Apple-iPhone-256GB-Silver-T-Mobile/product-reviews/Bo7RV52TRF/ref=cm_cr_dp_d_show_all_btm?/)' où nous allons extraire les avis des consommateurs vis-à-vis le produit Iphone Mobile.*

### *2-Etapes du projet :*

*#Commencer webscraping par cette commande pour créer dossier du projet scrapy startproject nom du projet (dans notre cas:[costumeReviews](#))*

*# créer spider (qui permet d'extraire les données à partir des pages web)  
C:\Users\PC\PycharmProjects\WebScraping>  
scrapy genspider <spidername>(ex :[reviewspider](#)) <your-link-here>(ex :  
[www.amazon.com](http://www.amazon.com))*

*#Remplir le fichier spider (reviewsspider.py) pour extraire data du site amazon  
(Voir le code qui est bien documenté)*

*#permet de remplir csv file avec data extrait  
([CostumeReviews](#)\CostumeReviews>scrapy crawl reviewspider -o  
scraped\_data2.csv (nom de fichier csv)*

*Je vous laisse le code qui est bien documenté pour comprendre les étapes de ce projet. Et si vous avez des questions n'hésitez pas de me contacter sur : « [meriem.nabil22@esi.ac.ma](mailto:meriem.nabil22@esi.ac.ma) »*