**Projet web scrapping**

* **Problématique**

Le monde de l’automobile est un marché ultra concurrentiel. Si on s’intéresse au marché de l’occasion, on remarque que les affaires se font de plus en plus rares alors que la concurrence ne fait que de croitre.

Comment pouvons-nous aider un revendeur auto à se démarquer afin de dénicher les perles rares avant tout le monde ?

* **Idée**

L’idée est de développer une API pour le revendeur dans laquelle il renseignerait certains filtres comme le type de véhicule qu’il recherche, le prix maximum ainsi que le kilométrage maximum.

Un scraping est alors effectué en fonction des filtres sur les sites La Centrale, l’Argus et potentiellement Le Bon Coin, l’utilisateur est notifié des annonces correspondantes.

Les sites sont rafraichis périodiquement et le scrapping est de nouveau effectué sur la première page de chaque site afin de récupérer de potentielles nouvelles annonces.

* **Aspect écologique**

Calcul du Crit’air (certificat qualité de l'air) pour chaque véhicule récupéré.

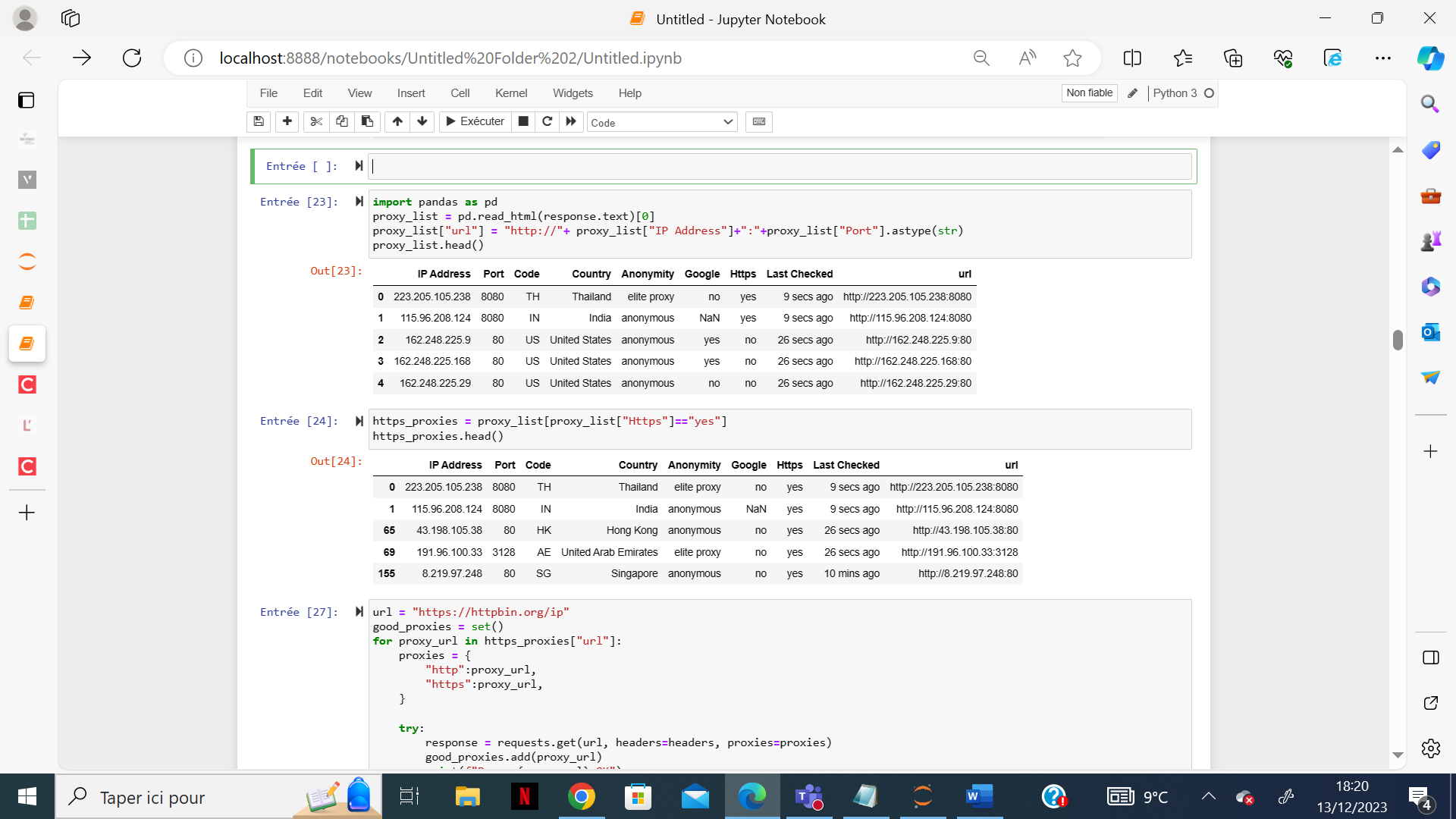
Une image contenant texte, logiciel, Icône d’ordinateur, Logiciel multimédia

Description générée automatiquement

Les **véhicules électriques** bénéficient d’office d’une vignette 0, qui leur est réservée. Les véhicules **hybrides rechargeables** ou à **gaz** écopent d’office d’une vignette Crit’Air 1.

* **Request ou Selenium ?**

Test du scraping avec Selenium en simulant les actions a effectué sur le site. Cependant, les sites détectent les mouvements du robot et bloquent ainsi toutes actions, même avec l’ajout de headers ou le changement d’adresse ip.



IP récupérés sur free proxy list

Nos requêtes étaient également détectées avec requests. Cependant, l’ajout de headers (notamment la modification du user-agent) a permis de d’effectuer les requêtes avec succès.

* **API**

L’interface sera crée grâce à la bibliothèque Tkinter de Python.