

Documentation du script de scraping

Prérequis

Pour exécuter le script, vous devez naviguer jusqu'à un répertoire spécifique sur votre système de fichiers. Une fois dans ce répertoire, vous devez ouvrir le script dans Visual Studio Code et l'exécuter manuellement. À l'avenir, une fonctionnalité sera ajoutée pour automatiser cette tâche à l'aide d'un bouton dans Streamlit en ligne.

Connexion à la base de données

Le script se connecte à une base de données Azure en utilisant des informations de connexion stockées dans un fichier `.env`. Ces informations comprennent le nom du serveur, le nom de la base de données, le nom d'utilisateur et le mot de passe. Ces informations sont chargées comme des variables d'environnement au début de l'exécution du script.

Scraping des données

Le script utilise Scrapy, un puissant framework de scraping en Python, pour récupérer les informations sur les pneus du site carter-cash. Le script génère dynamiquement les URL de départ en fonction des valeurs de largeur, hauteur et diamètre des pneus. Ces URL sont ensuite utilisées par Scrapy pour naviguer sur le site web et récupérer les informations.

Extraction des données

Les données sont extraites à partir des pages web en utilisant des sélecteurs spécifiques. Ces sélecteurs sont utilisés pour cibler et extraire des informations spécifiques sur chaque page. Les données extraites comprennent le prix, l'URL du produit, des informations générales, une description, la saisonnalité, le type de véhicule, la consommation, l'indice de pluie, le bruit, la largeur, la hauteur, le diamètre, la charge, la vitesse, si le pneu est runflat ou non, et une note. Ces données sont nettoyées et formatées avant d'être stockées dans un dictionnaire.

Insertion des données dans la base de données

Une fois les données extraites et formatées, elles sont insérées dans la base de données Azure. Les données sont insérées dans trois tables différentes : **Produit**, **Caracteristiques** et **Dimensions**. Chaque table stocke un ensemble spécifique d'informations. Par exemple, la table **Produit** stocke l'URL du produit, le prix, les informations générales, la description, la note et la date du scraping.

Optimisations

Le script utilise une bibliothèque spécifique pour interagir avec la base de données Azure, ce qui permet d'optimiser les performances de l'insertion des données. De plus, le script utilise une méthode spécifique pour s'assurer que toutes les modifications sont bien enregistrées dans la base de données.