Documentation du projet : Scraping et API FastAPI pour les annonces immobilières

Script de Scrapping :

**1.Étapes principales du script :**

1. **Accéder à la page des annonces** (http://www.tunisieannonce.com/AnnoncesImmobilier.asp)
2. **Scraper les annonces** :
   * Récupérer les informations sous forme de texte.
   * Extraire les éléments importants (région, type, prix, date, etc.).
3. **Sauvegarder les annonces** sous forme de **tableau** en utilisant **Pandas**.
4. **Exporter les données dans un fichier Excel** (annonces.xlsx).

2. **Explication du Code**

**1. Importation des bibliothèques nécessaires**

Le script commence par installer beautifulsoup4 (Pour **analyser et extraire** les données HTML de la page) et importer les **modules Python** requis tels que pandas et requets(Pour envoyer une **requête HTTP** et récupérer le contenu de la page) :

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, Police, capture d’écran

Le contenu généré par l’IA peut être incorrect.

**2. Envoi de la requête HTTP à la page des annonces**

Le script **envoie une requête** pour récupérer le contenu HTML de la page

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**3. Vérification du succès de la requête**

On s'assure que la requête **a bien réussi** (status\_code == 200 signifie que la page a été chargée correctement)



**4. Extraction des annonces**

On utilise **BeautifulSoup** pour **analyser le code HTML** et **trouver** les annonces

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**5. Traitement des annonces**

Chaque **annonce** est analysée **séparément** pour extraire les informations utiles

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**6. Stockage des annonces**

Les annonces sont **ajoutées** à une liste pour un traitement ultérieur

Une image contenant texte, capture d’écran, Police, conception

Le contenu généré par l’IA peut être incorrect.

**7. Conversion en tableau avec Pandas**

Une fois toutes les annonces collectées, elles sont **converties** en un **DataFrame** avec **Pandas**



**8. Exportation vers Excel**

Les données sont enregistrées dans un fichier **Excel**

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.



Fonctionnalité de l’API:

1. **Installation et Importation des bibliothèques**

Le script commence par installer et importer les bibliothèques nécessaires :

Une image contenant texte, capture d’écran, Police, conception

Le contenu généré par l’IA peut être incorrect.

2. **Initialisation de l'API**

Création d’une instance de l’application FastAPI :

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Le contenu généré par l’IA peut être incorrect.

3. **Fonction de Scraping**

La fonction scraper() récupère les annonces depuis **Tunisie-Annonce** :

* Envoie une **requête GET** au site Tunisie-Annonce
* Analyse le HTML avec **BeautifulSoup**
* Récupère les annonces et les stocke dans annonces\_data

Une image contenant texte, capture d’écran, logiciel, Police

Le contenu généré par l’IA peut être incorrect.

4. **Exécution du Serveur FastAPI dans Google Colab**

Colab ne supporte pas uvicorn.run() normalement du coup j’ai utiliser **nest\_asyncio** permet d’exécuter FastAPI dans un notebook :

Une image contenant texte, capture d’écran, Police, logiciel

Le contenu généré par l’IA peut être incorrect.

**5.Hébergement via Ngrok**

Ngrok permet d’accéder à l’API depuis une URL publique :

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, ligne, Police, capture d’écran

Le contenu généré par l’IA peut être incorrect.

https://16d6-35-222-104-157.ngrok-free.app/docs#/

Une image contenant texte, logiciel, Logiciel multimédia, Logiciel de graphisme

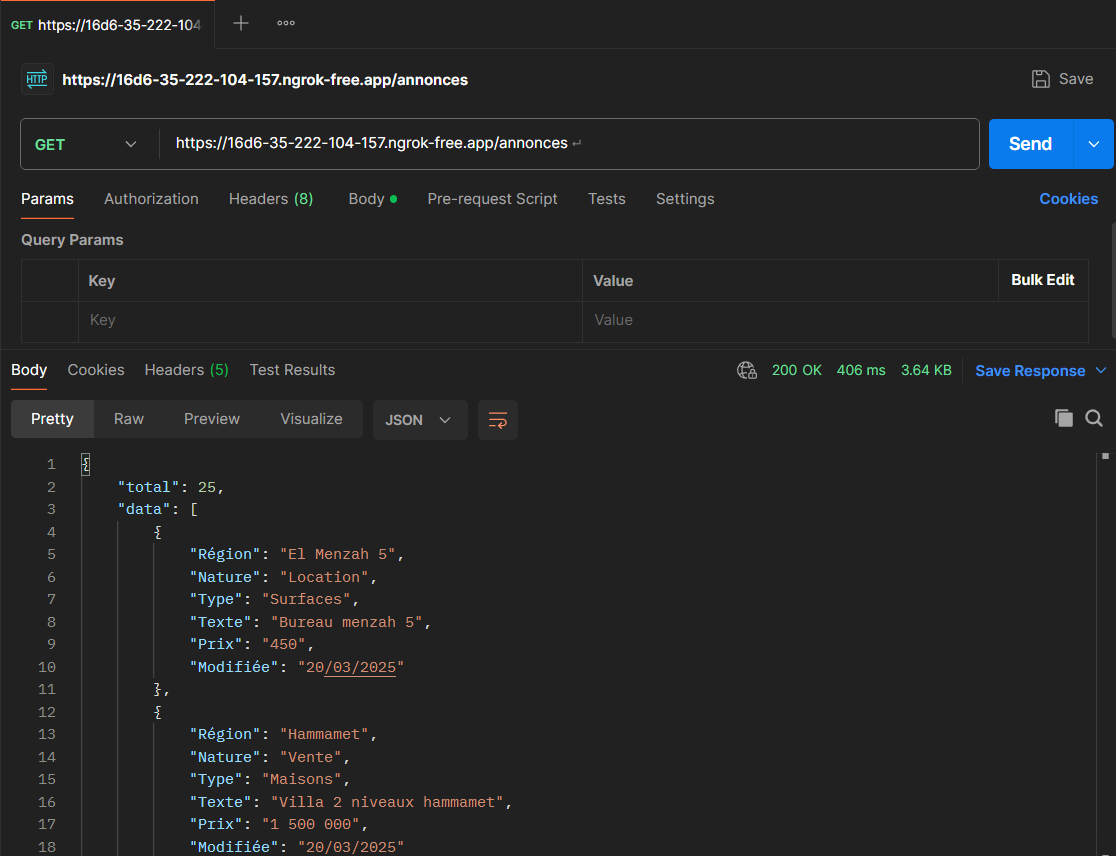
Le contenu généré par l’IA peut être incorrect.

<https://dashboard.ngrok.com/get-started/your-authtoken>

Exécution sur Postman :

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Le contenu généré par l’IA peut être incorrect.



✅ **Scraping automatique des annonces immobilières**   
✅ **API REST avec FastAPI pour récupérer les données**   
✅ **Documentation claire et détaillée**   
**✅ Possibilité d'hébergement via Ngrok (Colab)**

**✅Le code source sur un dépôt GitHub.**