

Université Chouaïb Doukkali

Ecole supérieure de technologie Sidi Bennour

Rapport de projet Python.

Développement d'une application desktop: Web scraping

Réalisateur:

- Elbahi Ayoub

Encadré par :

- M.Badi Youssef

Année universitaire : 2022-2023

Sommaire

Sommaire.....	2
1.Introduction.....	3
2.Développement d'une application Desktop Web Scraping.....	3
2.1.Definition.....	3
2.2.Les contraintes de réalisation.....	4
3.Développement de l'application.....	4
3.1.Outils de développement.....	4
3.2.Les bibliothèques utilisées.....	5
3.3.L'interface graphique.....	6
3.4.La barre de menu.....	7
3.5.Lancer une recherche.....	9
4.Programmation.....	9
5.Conclusion.....	10

Introduction

Ce rapport présente les résultats d'un projet de web scraping réalisé en utilisant les bibliothèques Python et BeautifulSoup. Le but de ce projet était d'extraire les données d'un site web cible et de les stocker dans un format exploitable pour une analyse ultérieure. Les étapes du processus de scraping, de la sélection de la source de données à l'analyse et à la présentation des résultats, sont décrites en détail dans les sections suivantes. Les défis rencontrés et les solutions apportées sont également discutés. Enfin, les conclusions et les suggestions pour des travaux futurs sont présentées. Le projet est à réaliser durant le semestre 5 en binôme ou monôme. J'ai donc décidé de le faire seul. Le travail est à rendre pour la fin du premier semestre.

2.Développement d'une application Desktop Web Scraping

2.1 Définition:

Le web scraping est un processus automatisé d'extraction de données à partir d'une source en ligne. Cela peut être utile pour collecter des informations sur des produits, des prix, des tendances, des statistiques, etc. Python est un langage de programmation populaire pour le web scraping en raison de sa simplicité et de sa puissance. La bibliothèque BeautifulSoup est un outil puissant pour extraire des données de pages web en HTML. Il permet de parser le code HTML d'une page et de l'analyser pour trouver les informations que l'on cherche. Avec BeautifulSoup, vous pouvez facilement trouver des informations spécifiques sur une page, telles que des balises, des classes et des id, et extraire les données associées. Lors du

développement d'un projet de web scraping avec Python et BeautifulSoup, vous pouvez définir les pages web que vous souhaitez extraire, télécharger le code HTML de ces pages et le parser avec BeautifulSoup pour trouver les informations que vous recherchez. Vous pouvez ensuite enregistrer ces données dans un fichier ou une base de données pour une analyse ultérieure. En résumé, un projet de web scraping avec Python et BeautifulSoup consiste à automatiser l'extraction de données à partir de pages web en utilisant une combinaison de Python et de BeautifulSoup pour analyser et extraire des informations utiles.

2.2 Les contraintes de réalisation:

Lors de la réalisation d'un projet de web scraping, plusieurs contraintes peuvent être rencontrées. Tout d'abord, il est important de prendre en compte les politiques de protection des données et de confidentialité des sites web. Certaines pages web peuvent interdire le scraping en utilisant des mécanismes tels que des protections anti-bots ou des restrictions IP. Il est donc important de vérifier les politiques de chaque site avant de commencer le scraping. De plus, la structure du code HTML d'une page peut varier considérablement d'un site à l'autre, ce qui peut rendre difficile l'extraction de données spécifiques. Cela peut nécessiter une analyse approfondie de la structure de la page pour déterminer comment extraire les informations désirées. En outre, le fait que les sites web puissent être mis à jour fréquemment peut également poser un défi pour le scraping. Si la structure de la page change, cela peut entraîner une modification du code utilisé pour extraire les données, ce qui peut entraîner une maintenance régulière du script de scraping. Enfin, il est important de prendre en compte les performances et la rapidité du script de scraping. Le scraping de grandes quantités de données peut nécessiter beaucoup de temps et peut être limité par la vitesse de la connexion Internet ou par les ressources système disponibles. Il est donc important de planifier et de tester soigneusement le script pour s'assurer qu'il fonctionne efficacement et de manière optimale.

3.Développement de l'application

3.1. Outils de développement

a) Langage de programmation

Pour développer cette application j'ai utilisé le langage de programmation Python. Python est un langage de programmation, dont la première version est sortie en 1991. Créé par Guido van Rossum. C'est un langage de programmation multi-paradigme. Il favorise la programmation impérative structurée, et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions ; il est ainsi similaire à Perl, Ruby, Scheme, Smalltalk et Tcl.

b) Plateforme de développement

Et j'ai utilisé comme plateforme de développement Visual Studio Code, il fournit divers outils pour le développement de logiciel, incluant les activités de codage logiciel (avec un environnement de développement intégré complet) ainsi que de la modélisation, de la conception, des tests, du reporting, etc. Son environnement de développement est conçu pour être polyvalent et prend en charge de nombreux langages de programmation.

3.2. Les bibliothèques utilisées

a) BeautifulSoup

J'ai utilisé BeautifulSoup pour extraire les données à partir de pages web. C'est une bibliothèque puissante pour le traitement du HTML qui facilite la localisation et l'extraction de données spécifiques sur les pages web. Il permet de parser facilement le code HTML d'une page et de l'analyser pour trouver les informations souhaitées. De plus, BeautifulSoup est compatible avec plusieurs langages de programmation, ce qui le rend polyvalent et facile à utiliser pour de nombreux projets de web scraping.

b) Tkinter

Et j'ai utilisé Tkinter comme plateforme de développement, c'est une bibliothèque de la bibliothèque standard de Python pour la création d'interfaces graphiques. Il fournit divers outils pour la création d'applications graphiques en utilisant Python, y compris des fenêtres, des boutons, des champs de saisie, des étiquettes et d'autres composants graphiques. Tkinter est facile à utiliser et permet de créer rapidement des interfaces graphiques attrayantes et fonctionnelles pour les applications Python.

c) Requests

La bibliothèque Requests est un outil puissant pour les développeurs Python. C'est une bibliothèque HTTP qui permet d'envoyer des requêtes HTTP à des serveurs web et d'obtenir

des réponses en utilisant Python. Elle offre une interface simple et conviviale pour effectuer des tâches courantes telles que l'envoi de requêtes GET et POST, la gestion des redirections et des erreurs, l'ajout de paramètres et d'entêtes HTTP, et la gestion des cookies. La bibliothèque Requests simplifie grandement le traitement des données provenant d'internet en fournissant une méthode élégante et concise pour envoyer et recevoir des requêtes HTTP. Cela en fait un choix populaire pour les projets de web scraping, de développement d'API et de toutes les applications nécessitant une interaction avec des serveurs web.

3.3) L'interface graphique

L'interface graphique de l'application doit être ergonomique, conviviale et facile à utiliser. Cette interface permettra à l'utilisateur de rechercher un produit sur le net. La figure suivante illustre la fenêtre de l'application au démarrage :



Figure 1:Fenêtre de l'application

Cette interface comprend plusieurs widget:

- **Entry box** : ce widget permet d'insérer le produit que l'utilisateur veut rechercher.
- **Website dropdown**: il permet de basculer entre les différents sites web à scraper.
- **Scrape button**: permet de lancer le processus de recherche.
- **Textbox** : permet d'afficher les résultats obtenus.
- **Progression bar** : ce widget permet d'afficher l'état de progression de l'opération.

3.4.La barre de menu:

cette barre contient trois menus :

File Help Contact us

Figure 2 : La barre de menu.

- a) **Menu Fichier** : ce menu contient les sous menus suivants:
- i) Exporter vers CSV: pour exporter les résultats sous un fichier .csv.
 - ii) Quitter : pour quitter l'application

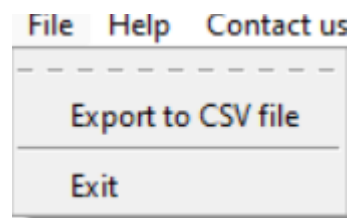


Figure 3 : le menu fichier.

b) **menu aide** : contient le guide d'utilisation qui s'ouvre dans une mini fenêtre.

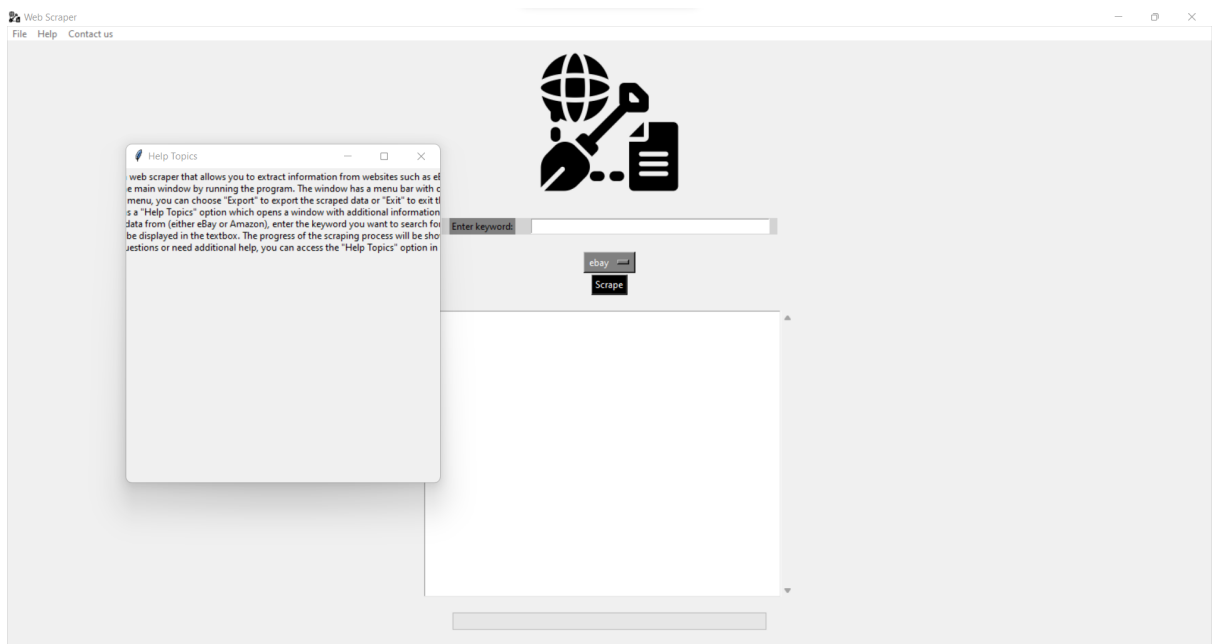


Figure 4 : Menu Help.

c) **Menu contactez-nous** : contient l'adresse mail du support.

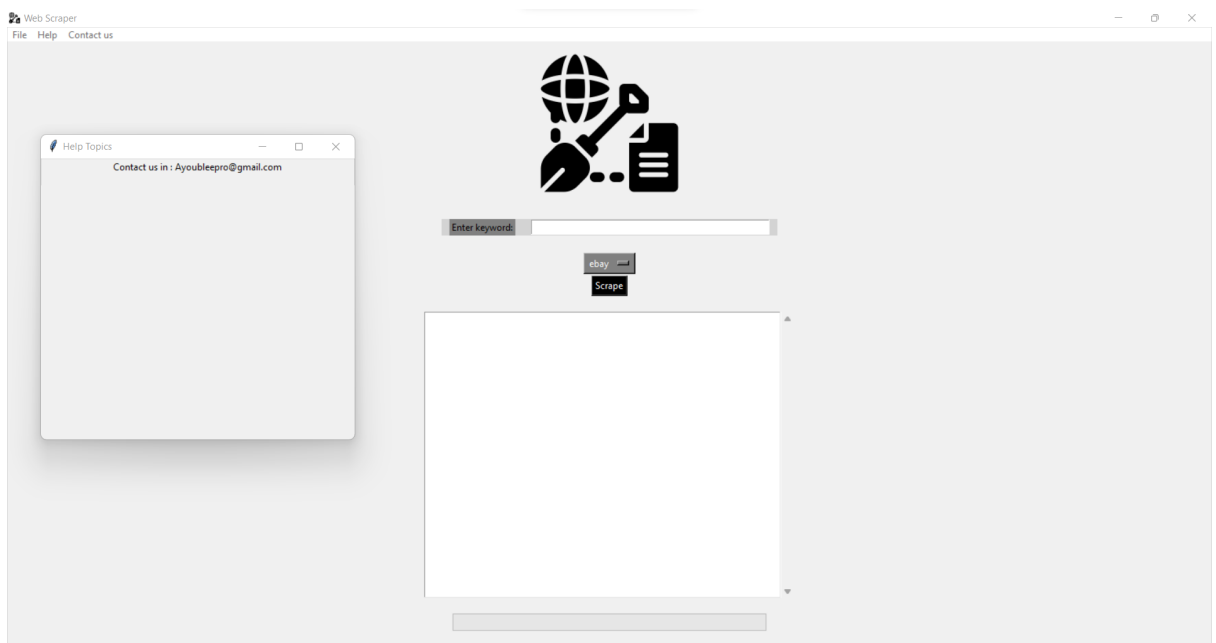


Figure 5 : Menu contact us.

3.5. Lancer une recherche

Pour lancer une recherche il faut suivre les étapes suivantes :

1. Insérez un mot-clé que vous voulez utiliser pour rechercher un produit.
2. Choisissez le site que vous voulez scraper en cliquant sur le dropdown.
3. Cliquez sur le bouton nommé "scrape" pour lancer l'opération de recherche.
4. Vous pouvez également exporter les résultats dans un fichier CSV en utilisant le menu **File > Export file**.

4. La programmation

The web scraping functionality is performed when the "scrape" function is called. It uses the requests library to make a GET request to the selected website's URL, which is built based on the keyword entered by the user and the website selected from the drop-down menu. The BeautifulSoup library is used to parse the HTML content returned from the GET request. The scraped product data is stored in a list of dictionaries, where each dictionary contains information about a single product (description, price, and link). The data is then displayed in a text box in the main window :

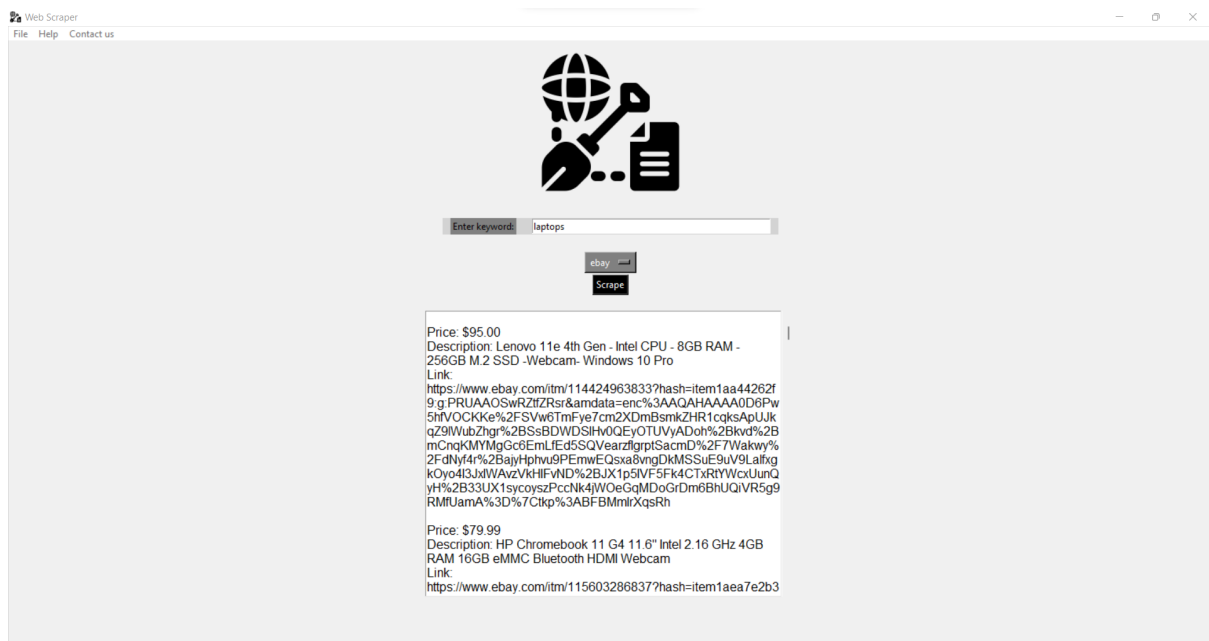
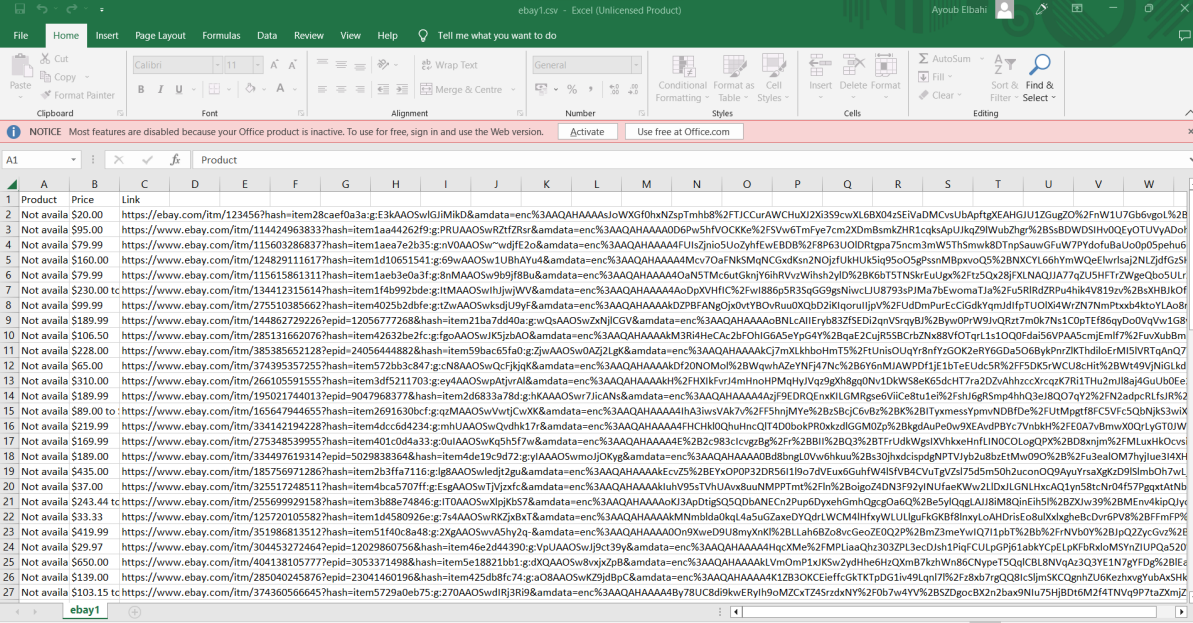


Figure 6 : Affichage des résultats.

il est totalement possible de stocker les résultats de recherche obtenues dans un fichier excel en utilisant l'option "Export to csv" on aura donc :



Product	Price	Link
Not availa	\$20.00	https://ebay.com/itm/1234567hash=item28cae0a3a:g:E3kAAOSwGiiMikD&amdata=enc%3AAQAHAAAsIoWQGI0hXNzpTmh82FTICCurAWCHuXJ2K6359cWxL6B8X045EiVaDMcVubApfKfXEAHGIU12GugZ0%2fW1U7Gb6vgoL%2B
Not availa	\$95.00	https://www.ebay.com/itm/1144249638337hash=item1aa4262f9:g:PRUAAOSwRztrZr&amdata=enc%3AAQAHAAAD06pw5HVOCKK%2f5Vw6tmFye7cm2XDmBsmkZHR1cqsApUlkqZ9IWubZhg%2B5sBDDWDSHwQOEYOTUvYyAaDol
Not availa	\$79.99	https://www.ebay.com/itm/115603286837hash=item1aea7e2b35:g:nV0AAOSw~vdjFE2o&amdata=enc%3AAQAHAAAD4FUISZpio5UoZyhfEwEBDB%2F8P63UOIdRTgaa75ncm3mW5THSmwk8DTnpSawwGfUw7PYdofuBaUo0p05pehu6
Not availa	\$160.00	https://www.ebay.com/itm/124829111617hash=item1d10651541:g:69wAAOSw1UBhAy4&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$79.99	https://www.ebay.com/itm/115615861311hash=item1aeb3e0a3f:g:8nMAOSw9b9jF8Bu&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$230.00	https://www.ebay.com/itm/134412315614hash=item1f4b992bde:g:ITMAAOSwHjWVW&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$99.99	https://www.ebay.com/itm/275510385662hash=item4025b2dfe:g:tZwAAOSwksdjU9f7&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$189.99	https://www.ebay.com/itm/1448627292267epid=12056777268&hash=item21ba7d4daa:g:wQaAAOSwZxNjCjGV&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$106.50	https://www.ebay.com/itm/285131662076hash=item4263be2fc:g:fgoAAOSwK5jzAO&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$228.00	https://www.ebay.com/itm/3853856521287epid=24056444882&hash=item59bac65fa0:g:ZjwAAOSw0A2JLg&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$65.00	https://www.ebay.com/itm/374395357255hash=item572bb3c847:g:cN8AAOSwQcFjKjK&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$310.00	https://www.ebay.com/itm/266105591555hash=item57d5211703:g:ey4AAOSwPajvriAI&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$189.99	https://www.ebay.com/itm/1950217440137epid=90479683778hash=item206833a78d:g:kAAASw7JicAN&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$99.00	https://www.ebay.com/itm/165647944655hash=item26916300cf:g:qMAAOSwVwVjcw&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$219.99	https://www.ebay.com/itm/334142194287hash=item4dc64234:g:mhUAAOSwQvdkh17&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$169.99	https://www.ebay.com/itm/3744548539955hash=item401c0d4a33:g:0uIAAOSwKq5h57w&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$189.00	https://www.ebay.com/itm/33449376193147epid=5029838364&hash=item4de19c9d72:g:ylIAAOSwvnoJOKg&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$435.00	https://www.ebay.com/itm/1855769712867hash=item2b3ffa7116:g:lg8AAOSwlejd2gu&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$37.00	https://www.ebay.com/itm/125517248511hash=item4bca5707ff:g:Eg8AAOSwTjVjzfc&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$243.44	https://www.ebay.com/itm/255699929158hash=item3d88e74846:g:IT0AAOSwXlpjKb57&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$33.33	https://www.ebay.com/itm/125720105582hash=item1d4580926e:g:7s4AAOSwRkZjXB7&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$419.99	https://www.ebay.com/itm/351986813512hash=item5140c8a48:g:2XgAAOSwvA5hy2q&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$29.97	https://www.ebay.com/itm/304453274647epid=12029860756&hash=item46e2d44390:g:PvUAAOSwJ9c3t39y&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$650.00	https://www.ebay.com/itm/4041381057777epid=3053371498&hash=item5e18821bb1:g:dXQAAOSw8vjkpB&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$139.00	https://www.ebay.com/itm/2850402458767epid=23041460196&hash=item425db8fc74:g:aO8AAOSwK29jBpC&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf
Not availa	\$103.15	https://www.ebay.com/itm/374360566645hash=item5729a0eb75:g:270AAOSwdlRj3R&amdata=enc%3AAQAHAAAD4A0N5TMC6utGknj76iHrVvzWihsh2yID%2B86bT5TNSKrUoUg%2f2fz5Q28JFLXNAQJIA77QZUHFTHZwgeQbo5UoUf

Figure 7 : Fichier Excel.

5.Conclusion

En conclusion, le développement d'un web scraper en utilisant Python a permis d'obtenir une solution efficace et puissante pour extraire des données à partir de sites web. La flexibilité de Python en matière de développement de scripts a facilité la création d'un scraper robuste capable de gérer les erreurs et les exceptions, tout en offrant une grande vitesse et une performance optimale. L'utilisation de bibliothèques dédiées telles que BeautifulSoup et Requests a également simplifié la tâche en fournissant des fonctionnalités prédéfinies pour manipuler et analyser les données récupérées. En fin de compte, ce rapport de développement démontre la puissance et la flexibilité de Python pour la création de web scraper performants pour les entreprises et les organisations.