

[← Mon parcours](#) PROJET VALIDÉ

Utilisez les bases de Python pour l'analyse de marché

[Mission](#) [Cours](#) [Ressources](#) [Évaluation](#) 60 heures

Mis à jour le lundi 13 mars 2023

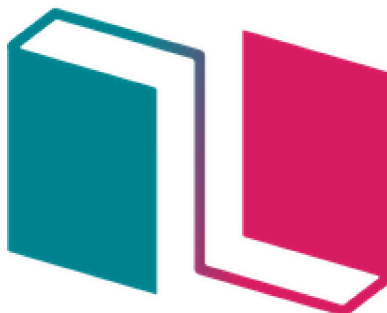
Si vous avez commencé ce projet avant le 28/09/2022, vous avez débuté votre travail sur **ce projet archivé**. Vous pouvez continuer l'existant ou recommencer avec ce nouveau projet, dont les modifications s'alignent avec les retours étudiants.

Avant de démarrer votre travail sur ce projet, nous vous conseillons de:

1. Lire le scénario en entier, chaque section du projet ainsi que les documents fournis.
2. Consulter le **guide des étapes clés** avec des recommandations et des ressources pour organiser votre travail.
3. Préparer une liste de questions pour votre première session de mentorat.

Scénario

Vous êtes analyste marketing chez Books Online, une importante librairie en ligne spécialisée dans les livres d'occasion. Dans le cadre de vos fonctions, vous essayez de suivre manuellement les prix des livres d'occasion sur les sites web de vos concurrents, mais cela représente trop de travail et vous n'arrivez pas à y faire face : il y a trop de livres et trop de librairies en ligne ! Vous et votre équipe avez décidé d'automatiser cette tâche laborieuse via un programme (un scraper) développé en Python, capable d'extraire les informations tarifaires d'autres librairies en ligne.



Books Online

Sam, votre responsable d'équipe, vous a chargé de développer une version bêta de ce système pour suivre les prix des livres chez [Books to Scrape](#), un revendeur de livres en ligne. En pratique, dans cette version bêta, votre programme n'effectuera pas une véritable surveillance en temps réel des prix sur la durée. Il s'agira simplement d'une application exécutable à la demande visant à récupérer les prix au moment de son exécution.

Sam vous a envoyé l'e-mail suivant :

Objet : Programme d'extraction des prix

À : Vous

De : Sam

Bonjour !

J'espère que vous pourrez m'aider à créer un système de surveillance des prix. Pour élaborer une version bêta du système limitée à un seul revendeur, Books to Scrape, le mieux est probablement de suivre les étapes que j'ai définies dans le fichier des exigences ci-joint.

Lorsque vous aurez terminé, envoyez-moi un lien vers votre **repository GitHub** et un fichier compressé des données qu'il génère. Avec le code, le repo doit inclure un requirements.txt et un README.md complété afin que j'exécute le code avec succès et produise des données ! Le repo ne doit pas inclure les données et images extraites.

Après avoir terminé le code, envoie-moi un **fichier ZIP des données générées**. Assurez-vous d'organiser toutes les données et images que vous avez extraites de manière simple.

Pouvez-vous également m'envoyer un **mail** décrivant comment nous pourrions utiliser le code pour établir un pipeline ETL (de l'anglais **E**xtract, **T**ransform, **L**oad, signifiant extraire, transformer, charger) ? Cela sera utile pour le montrer à mon responsable.

Cordialement,

Sam

Responsable d'équipe

Books Online

Pièce-joint :

[**Exigences du système de surveillance des prix**](#)

Ces besoins étant clarifiés, vous êtes prêt à mettre votre maîtrise de Python toute neuve au service de votre équipe !

Pour rappel, afin de vous aider à réaliser ce projet, vous pouvez consulter [**ce guide d'étapes clés**](#), qui contient des recommandations et des ressources pour organiser votre travail.

Livrables

1. Un document TXT ou PDF contenant le lien vers le **repository GitHub** public qui doit contenir les éléments suivants:
 - l'ensemble de votre code d'application ;

- le fichier requirements.txt, mais pas l'environnement virtuel lui-même ;
 - un fichier README.md expliquant comment créer et activer l'environnement virtuel, puis exécuter le code d'application ;
 - les données/images extraites ne doivent pas faire partie du repository lui-même.
2. Un fichier compressé ZIP contenant toute la **data** : les **données extraites et les images associées** dans un format ou une structure facile à suivre.
 3. Un **mail** (pas plus d'une page) **au responsable d'équipe**, Sam, décrivant comment l'application permet d'établir un pipeline ETL, au format PDF.

Pour faciliter votre passage devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre_du_projet_nom_prénom**", tous les livrables du projet comme suit : **Nom_Prénom_n° du livrable_nom du livrable__date de démarrage du projet**. Cela donnera :

- *Nom_Prénom_1_repo_mmaaaa*
- *Nom_Prénom_2_data_images_mmaaaa*
- *Nom_Prénom_3_mail_mmaaaa*

Par exemple, le premier livrable peut être nommé comme suit : *Dupont_Jean_1_repo_012022*

Soutenance

Durant la présentation orale, l'évaluateur interprétera le rôle de Sam, votre responsable d'équipe. La soutenance est structurée de la manière suivante :

- **Présentation des livrables (15 minutes)**
 - Présentez l'application en décrivant le processus utilisé pour l'écrire et en mettant en évidence les sections extraction, transformation et chargement du code.
 - Faites à votre évaluateur une démonstration de l'application en l'exécutant « en direct ».
 - Présentez des idées d'améliorations futures du code.
- **Discussion (10 minutes)**
 - L'évaluateur jouera le rôle de votre responsable d'équipe. Il vous challengera sur votre méthode et vos livrables. Par exemple :
 - Comment avez-vous identifié les champs de données pertinents que vous avez extraits de la page Produit ?
 - Comment vous êtes-vous assuré que l'environnement virtuel et les données extraites ne sont pas stockés dans le repository ?
 - Pouvez-vous expliquer comment votre programme permet d'établir un pipeline ETL (extract, transform, load) ?
- **Debrief (5 minutes)**
 - À la fin de la soutenance, l'évaluateur arrêtera de jouer le rôle de Sam, votre responsable d'équipe pour vous permettre de débriefer ensemble.

Votre présentation devrait durer 15 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, les présentations en dessous de 10 minutes ou au-dessus de 20 minutes peuvent être refusées.

Compétence transversale

En plus des compétences spécialisées énumérées ci-dessous, concentrez-vous sur le développement et la démonstration de la **communication** en tant que compétence transversale (ou "soft skill"). Elle sera essentielle pour votre réussite dans ce projet et les suivants, ainsi que dans votre future carrière. N'hésitez pas à demander à votre mentor un retour au cours de la réalisation de votre projet.

Compétences évaluées



Gérer les données à l'aide du processus ETL



Utiliser le contrôle de version avec Git et GitHub



Appliquer les bases de la programmation en Python

OPENCCLASSROOMS



OPPORTUNITÉS



AIDE



POUR LES ENTREPRISES



EN PLUS



Français



Télécharger dans
l'App Store