

# Accueil

🕒 100 heures

Mis à jour le mardi 12 mars 2024



Ce projet a été mis à jour le 7 juin 2023, si vous avez déjà commencé à travailler sur ce projet, vous trouverez ici **l'ancien énoncé**. Sinon vous pouvez commencer le scénario ci-dessous.

Une grille d'auto-évaluation de votre travail est disponible dans la partie "livrable" de ce projet.

L'entreprise Lapage était originellement une librairie physique avec plusieurs points de vente. Mais devant le succès de certains de ses produits et l'engouement de ses clients, elle a décidé depuis 2 ans d'ouvrir un site de vente en ligne.



Logo de Lapage

La structure a besoin d'aide pour mieux comprendre ses données, c'est pourquoi elle vous a recruté. Vous intervenez en tant que Data Analyst afin de faire le point sur l'activité. L'équipe Marketing que vous intégrez se compose de :

- Antoine, Analyste d'Affaires.

Annabelle vous a déjà expliqué le contexte et les attendus de la mission oralement mais elle vous a fait un récapitulatif par email :

**Objet** : Bienvenue dans l'équipe - briefing de mission

**De** : Annabelle

**À** : Moi

Bonjour,

Bon courage pour ta prise de poste. J'espère que tu te pliras parmi nous.

Comme tu l'as certainement compris, nous avons ouvert il y a deux ans de cela notre librairie en ligne et nous aimerions maintenant faire le point sur les différents indicateurs et chiffres clés de l'entreprise.

Cela nous permettra de décider de la marche à suivre, par exemple décider si nous devons créer certaines offres, adapter certains prix ou cibler un certain type de clientèle etc.

Cette analyse va être découpée en deux parties :

- Une analyse des différents indicateurs de vente, qui a été demandée directement par Antoine. Je t'ai mis en pièce jointe l'e-mail qu'il m'a envoyé résumant ses demandes.
- Une analyse plus ciblée sur les clients : l'objectif est de comprendre le comportement de nos clients en ligne, pour pouvoir ensuite comparer avec la connaissance acquise via nos librairies physiques. Pour cela, je t'invite à voir avec Julie, qui te précisera directement ses demandes à ce niveau.

La base de données est en pièce jointe. N'hésite pas si tu as des questions, je me tiens à ta disposition.

Excellente journée à toi !

**Annabelle**

**Pièces jointes :**

- [Mail forwardé d'Antoine](#)
- [Fichier ZIP contenant les bases](#)

Après avoir consulté le mail d'Antoine, vous prenez directement contact avec Julie sur Teams.

chargé d'analyser les données de nos ventes en ligne. Annabelle m'a dit que tu avais des demandes spécifiques concernant l'analyse des comportements clients

**Julie** 10:16

Salut !

10:17

Bienvenue dans l'équipe !

J'allais justement prendre contact avec toi car je vais avoir besoin que tu regardes certaines corrélations. Nous souhaitons mieux comprendre le comportement de nos clients.

Bien sûr. Tu as déjà une idée précise de ce que tu veux que je regarde ?

**Julie** 10:18

Oui ! J'aimerais que tu regardes en particulier 5 corrélations :

10:18

le lien entre le genre d'un client et les catégories des livres achetés,

le lien entre l'âge des clients et le montant total des achats,

le lien entre l'âge des clients et la fréquence d'achat,

le lien entre l'âge des clients et la taille du panier moyen,

et pour finir le lien entre l'âge des clients et la catégorie des livres achetés.

Ça marche, je m'occupe de ça et je reviens vers toi si j'ai des questions ou pour te présenter les résultats !

Voici la version **pdf** de la conversation.

Pour vous aider dans la réalisation des tests statistiques, vous pouvez consulter le **tableau récapitulatif d'XLSTAT**.

Vous êtes prêt ? Votre mission démarre :

- Coupez dès à présent toutes les sources de distraction (téléphone, messagerie instantanée, mails..)
- Évitez les situations de multi tâches (n'écoutez pas un podcast ou les informations en travaillant)
- Préparez votre environnement de travail (onglet, documents, raccourcis...) et démarrez la mission en téléchargeant les fichiers.

Bon courage !

Dans ce projet, le fichier qui vous est fourni est nettoyé, cependant si vous souhaitez vous entraîner vous pouvez décider de partir de **ce fichier-ci** et pourrez ensuite le comparer avec le fichier donné dans le projet afin d'être sûr de faire vos analyses sur les bonnes bases.

Il y a plusieurs façons de nettoyer les données, l'important c'est la logique et la méthodologie que vous avez choisies. N'hésitez pas à en parler avec votre mentor.

Dans ce projet, vous êtes libre d'utiliser le langage de programmation de votre choix entre R et Python pour mener à bien votre analyse. Vous pouvez utiliser Jupyter (pour R et pour Python) ou R Markdown (pour R).



## Livrables

Vous pouvez vérifier la qualité de votre travail en utilisant cette **check-list d'auto-évaluation**.

1. Le ou les notebooks R Markdown ou Jupyter contenant les traitements et les différentes analyses effectués pour répondre aux demandes d'Antoine et Julie.

Pour faciliter votre passage devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre\_du\_projet\_nom\_prénom**", votre livrable nommé comme suit : **Nom\_Prénom\_n° du livrable\_nom du livrable\_date de démarrage du projet**. Cela donnera :

- *Nom\_Prénom\_1\_notebook\_mmaaaa*

Par exemple, votre livrable peut être nommé comme suit :  
*Dupont\_Jean\_1\_notebook\_012023*.



## Soutenance

Pour cette soutenance, nous vous conseillons de rédiger un support de présentation mais celui-ci n'est pas obligatoire.

Durant la présentation orale, votre évaluateur jouera le rôle d'Annabelle à qui vous présentez vos résultats :

◦ Analyse des corrélations

- **Discussion (10 minutes)**

- L'évaluateur vous challengera sur vos choix.

- **Débriefing (5 minutes)**

- Votre évaluateur reprend son rôle de mentor pour que vous puissiez 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.

## Compétences évaluées



Réaliser un test statistique



Réaliser une analyse bivariée pour interpréter des données



Analyser des séries temporelles

Suivant