Scénario

Aujourd’hui est un grand jour, vous commencez votre mission en tant que data analyst freelance chez BottleNeck, un marchand de vin très prestigieux. Votre manager sur cette mission (Laurent) vous accueille chaleureusement et vous propose de partager un petit café avec le reste de l’équipe du service Numérique. L’ambiance est bonne, et vous voilà déjà parfaitement intégré dans cette équipe détendue mais professionnelle.

[](https://user.oc-static.com/upload/2020/11/23/16061348800222_Capture%20d%E2%80%99e%CC%81cran%202020-11-23%20a%CC%80%2013.33.23.png)Logo de Bottleneck

Après les premières présentations, effectuées dans une ambiance des plus conviviales, Laurent vous explique les enjeux de votre première mission :

*“Actuellement, pour gérer nos ressources, nos clients, etc., on utilise un ERP qui n’est absolument pas relié à notre site de vente en ligne. Pour être tout à fait honnête, les outils en place sont vraiment artisanaux et dans ces conditions, la gestion des stocks est vraiment complexe et notre visibilité en termes d’analyse des ventes sur le Net est vraiment réduite, car très peu de personnes ont accès au back-office. En attendant une solution plus centralisée, un rapprochement entre les 2 bases, même manuel, pourrait être très utile…*

*Ta première mission se passe en 3 points.*

*Premièrement, j’ai besoin que tu rapproches deux exports : un export de l’ERP contenant les références produit, leur prix de vente et leur état de stock, et un export d’une table de l’outil de CMS contenant les informations des produits commercialisés en ligne (nom, description, nombre de ventes...).*

*L’export issu de la boutique en ligne contient le nombre de ventes pour chaque produit depuis sa mise en ligne, il ne permet pas d’analyser l'évolution des ventes dans le temps.*

*Je vais t’envoyer un mail dès la fin de notre entretien avec ces 2 exports en pièce jointe.*

*En plus de ces 2 exports, tu vas bénéficier d’une aide précieuse car Sylvie, notre ancienne stagiaire, a réalisé un travail de fourmi. Elle a créé un tableau Excel qui permet d’établir le lien entre la référence du produit dans l’ERP (product\_id) et la référence du même produit dans la base de la boutique en ligne (SKU).*

*Deuxièmement, une fois le rapprochement effectué, je souhaiterais avoir le chiffre d’affaires par produit, ainsi que le total du chiffre d’affaires réalisé en ligne.*

*Troisièmement et pour finir, je me demande s’il n’y a pas eu des erreurs de saisie dans certains prix des produits. J'aimerais que tu effectues une analyse sur cette variable afin de détecter d’éventuelles valeurs aberrantes, de les lister et d’en faire une représentation graphique pour plus de lisibilité.*

*Nous voudrions présenter tes résultats lors de la prochaine réunion de COPIL. Cela permettrait de montrer nos progrès. Je peux t’épargner le travail d’une présentation par slides, il me faut juste un notebook que tu présenteras à l’assemblée pour expliquer ta démarche. Tu peux utiliser R ou Python comme tu préfères, nous n'avons pas de préférence de notre côté.”*

Après vous avoir méticuleusement délivré ces consignes, Laurent vous accompagne à votre poste de travail.

Juste avant de vous laisser prendre votre mission à bras-le-corps, il vous réitère qu’il va vous faire suivre les exports par mail, ainsi que la table de liaison de Sylvie.

Le mail des exports de tables :

**Objet**: Exports tables  
**De**: Laurent  
**À**: Moi

Re,

Voici les 2 exports dont nous avons parlé ce matin. Tu as le fichier issu de l’ERP (erp.xlsx) et le fichier de la table produit de notre plateforme de vente en ligne (web.xlsx). Je te laisse prendre connaissance de ces éléments.

N’hésite pas à me solliciter si tu as des questions.

Cordialement,  
**Laurent**

**Pièces jointes** :

* [erp.xls](https://s3-eu-west-1.amazonaws.com/static.oc-static.com/prod/courses/files/parcours-data-analyst/P5+ERP+vin/erp.xlsx)
* [web.xlsx](https://s3-eu-west-1.amazonaws.com/static.oc-static.com/prod/courses/files/parcours-data-analyst/P5+ERP+vin/web.xlsx)

Le mail du tableau de liaisons :

**Objet** : Fwd: Table de liaison  
**De**: Laurent  
**À**: Moi

---------- Forwarded message ---------

**Objet**: Table de liaison  
**De** : Sylvie  
**À** : Laurent

Bonjour Laurent,

Comme promis, voici en pièce jointe le fichier Excel qui liste les product\_id de l’ERP avec leur référence côté Web. La liste des product\_id est exhaustive, mais pour les références côté Web, j’en suis moins sûre... J’ai peiné à rapprocher certaines références.

Je suis désolée, mais je me rends compte à l’instant que j’ai également mal nommé la colonne dans le fichier Excel. En fait, la colonne id\_web dans mon fichier correspond au SKU des produits dans la boutique en ligne.

Je profite également de ce mail pour te remercier de m’avoir permis de faire mon stage avec toi. C’était une expérience formidable, et j’ai vraiment apprécié ton aide bienveillante.

Merci.  
À très bientôt.

Cordialement,  
**Sylvie**

**Pièce jointe** : [liaison.xlsx](https://s3-eu-west-1.amazonaws.com/static.oc-static.com/prod/courses/files/parcours-data-analyst/P5+ERP+vin/liaison.xlsx)

Cette fois-ci, vous avez toutes les cartes en main pour démarrer votre mission.

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 choix Jupyter (pour R et pour Python) ou R Markdown (pour R).

Livrables

* Le notebook R Markdown ou Jupyter.

Soutenance (30 minutes)

Pendant la soutenance, l’évaluateur jouera le rôle de Laurent, votre manager, à qui vous présentez votre notebook.

* Pendant **20 minutes** :
  + Vous expliquerez l’ensemble de votre démarche. Vous déroulerez votre notebook au mentor examinateur, en détaillant bien les actions entreprises pour obtenir un jeu de données final cohérent et complet.
  + Vous expliquerez également votre analyse univariée et vos conclusions.
* Pendant **10 minutes** : séance de questions-réponses.

La soutenance se terminera par un temps de débriefing avec l’évaluateur.

**Compétences évaluées**

 Classifier différents types de données

 Gérer les erreurs et les incohérences présentes sur des données stockées

 Réaliser une analyse univariée pour interpréter des données