

Cette mission suit un scénario de projet professionnel.

Aujourd’hui est un grand jour, vous commencez votre mission en tant qu’analyste chez BottleNeck, un marchand de vin prestigieux.

Nicolas, votre manager sur cette mission vous accueille chaleureusement et vous propose de partager un café avec le reste de l’équipe du service. L’ambiance est bonne et vous voilà déjà parfaitement intégré dans cette équipe détendue mais professionnelle.



Vous recevez un mail de Nicolas, votre manager.

Hello,

Suite à notre discussion de tout à l’heure au café, je souhaite te briefer sur la mission. Comme tu vas vite t’en rendre compte, nos outils sont artisanaux, nous avons du mal à analyser nos données et la gestion des stocks est complexe.

Ta mission se découpera en 2 phases :

- **Phase 1** : Agréger les différents fichiers afin de pouvoir exploiter les données.
- **Phase 2** : Analyser les données pour le CODIR.
  - Tu viendras faire la présentation de tes analyses au comité de direction (j’en profiterai pour te présenter à l’équipe de direction de l’entreprise).
  - Aline bien la présentation de tes analyses avec leurs préoccupations et sois synthétique !

Actuellement nous utilisons différents exports pour l’analyse :

- L’extraction de l'**ERP** (référence produit, prix et l’état du stock)

- L'extraction de notre **site Web** (SKU, quantités vendues, description des produits, etc.)
- Une **table de liaison** qui permet de lier les références entre la base de données Wordpress et l'extraction de l'ERP de l'entreprise.
  - Les numéros des références ne correspondent pas entre les deux outils.
  - Notre stagiaire a mis à jour cette liste pour toi avec les nouveaux produits.

Pour la **phase 1** :

Tu peux le faire avec un Notebook Python ou en R. Il faudra commencer par :

- rapprocher l'extraction de la BDD avec la base de données ERP via la table de liaison ;
- identifier les erreurs dans les données (je pense que tu peux en trouver au moins 8 : erreurs de saisie, de type, de calcul, de jointure, etc.) ;
- proposer des solutions pour améliorer les données dans nos systèmes.

Pour la **phase 2**, voici les analyses à faire :

- Calculer le chiffre d'affaires :
  - par produit ;
  - le total général.
- Travailler sur les tops références, le 20/80, etc.
- Vérifier les erreurs de saisie en détectant des potentielles valeurs aberrantes (tu peux utiliser un Z-Score ou l'écart interquartile).
- Extraire les valeurs aberrantes dans les prix (tu peux utiliser un boxplot pour la représentation) puis conclure si nous avons des erreurs de prix.
- Analyser l'état, les taux de marge, la rotation des stocks ou le nombre de mois de stock.

Pour finir, je souhaite savoir s'il y a une **corrélation** entre certaines données quantitatives (prix, prix d'achat, stock, ventes, prix HT et taux de marge, etc.). Est-ce que tu peux me faire cette analyse également ?

Je te joins le notebook sur lequel j'avais commencé à travailler, mais je n'ai jamais eu le temps de le terminer.

Si tu vois d'autres analyses pertinentes, n'hésite surtout pas !

Nicolas  
Responsable vente chez Bottleneck

**PS** : N'oublie pas que nous souhaitons vraiment comprendre les différentes erreurs que tu as pu rencontrer dans ton analyse. Nous allons utiliser ton rapport comme point de départ du futur projet de data visualisation. Tu peux, par exemple, nous faire un récapitulatif de toutes les corrections que nous allons devoir mettre en place pour avoir une base de données propre en présentant la formalisation du processus de ton travail et en justifiant tes choix de la conformité RGPD.

**PS2** : Concernant les données, c'est une extraction au 31 octobre et pour les ventes c'est du 1 octobre au 31 octobre.