

Test technique Data Engineer

On considère une application nommée *MyShopAnalytics* qui vise à aider le gérant d'un magasin en ligne à recueillir des informations statistiques concernant ses ventes.

Afin de répondre aux questions suivantes, vous avez à votre disposition une API REST ainsi qu'un dataset.

L'API REST donne accès à des informations sur les items et clients.

Le dataset au format CSV contient l'historique des transactions effectuées dans le magasin.

Votre tâche est de concevoir et implémenter une application qui exploite les données des clients, des items et des transactions afin de répondre aux fonctionnalités suivantes via une API (REST ou gRPC):

- Les **N items** les **plus vendus**.
- Les **N clients** qui **fréquentent* le plus** ce magasin.
- La **tranche d'âge** des clients qui **fréquentent* le plus** ce magasin parmi ces **intervalles (<18 ; 18 - 29 ; 30 - 55 ; > 55)** sur une **période temporelle**.
- Le **total dépensé (en euros)** pour un **genre**, une **période temporelle** et une **tranche d'âge**.
- Les **ventes (en nombre)** par **heure et jour de la semaine**.

Setup:

- Vous allez avoir besoin de Python 3.
- L'utilisation d'un environnement virtuel dédié est fortement recommandé.
- Afin de lancer l'API fournie, exécuter: `./start-api.sh` dans le dossier api.

N.B:

- L'utilisation de librairies Python est autorisée.
- Le projet doit être poussé sur un dépôt Github ou Gitlab **privé**.
- Ne pas pousser tout le projet en un unique commit. L'historique Git doit révéler votre progression pour aboutir au résultat final.
- Une grande attention sera accordée à la qualité du code livré et au respect des bonnes pratiques, dans un objectif de mise en production (ex. tests unitaires, maintenabilité...).

Bon courage,

L'équipe **InsideBoard**

* : Fréquenter = Nombre de jours distincts avec au moins une transaction.