

# Projet Fin période SAS

## Développeur Data

### Contexte du projet :

Ce projet vise à développer un pipeline de nettoyage de données utilisant Python, en se concentrant sur un ensemble de données d'e-commerce (Vente en ligne) de type CSV.

Les données seront nettoyées et ensuite stockées dans une base de données SQL pour faciliter leur utilisation ultérieure.

### Taches a réalisé :

Exploration et analyse des données : Utilisez des bibliothèques Python comme Pandas pour explorer les ensembles de données et identifier les problèmes de qualité des données tels que les valeurs manquantes, les doublons, les erreurs de formatage, etc.

Nettoyage des données : Appliquez des techniques de nettoyage de données pour résoudre les problèmes identifiés lors de l'étape précédente. Cela peut inclure le remplacement des valeurs manquantes, la suppression des doublons, la conversion des formats de données, etc.

Création de la base de données SQL : Utilisez une bibliothèque Python compatible avec SQL pour créer une base de données SQL et définir la structure des tables qui stockeront les données nettoyées. Assurez-vous de choisir les types de données appropriés pour chaque colonne.

**Quelle est la taille totale du Dataset avant et après le nettoyage des données ?**

**Y'a-il des valeurs manquantes dans le Dataset ? Si oui, comment les gérer ?**

**Les valeurs aberrantes ou les valeurs extrêmes doivent-elles être corrigées ou supprimées ? comment peut on faire**

**Quelle sont les clients qui ont passées plus de commande sur notre site ?**

**Quelle est le plus haut montant dépensé par un seul client ?**

**Quelles est la description des produits les plus vendus ou acheté par les clients dans notre site ?**

**Quel est le nombre moyen de produits vendus par jour, mois ?**

**Quelle est la répartition géographique des ventes ?**

**Quelle est la tendance des ventes au fil du temps, tracer un graphique de la croissance des ventes ?**

**Quelle est la saisonnalité des ventes ?**

**Quels sont les moments du jour (matin, après-midi, soir) que les clients achètent le plus ?**

**Quelle est la distribution des prix des produits ?**

**Après l'insertion des données nettoyées dans la base de données SQL, Créer des requêtes afin de vérifier des informations que vous voyez utiles du Dataset.**