

TP : Introduction Pratique à la Science des Données et au Processus de Science de données

0. Créer un nouveau Notebook dans Jupyter

1. Chargement et aperçu du jeu de données

- Importer les bibliothèques de base : pandas, numpy, et matplotlib ?
- Charger le fichier CSV **ventes_produits.csv** dans un DataFrame pandas, télécharger le fichier à partir l'URL <https://elearning-facsci.univ-annaba.dz/mod/folder/view.php?id=25203>
- Afficher le contenu de fichier **ventes_produits.csv**. Affichez les cinq premières lignes du tableau. Que remarquez-vous ?

2. Objectifs de recherche :

- Quels sont les objectifs potentiels de cette analyse de données ?
- Formulez quelques questions de recherche que vous pourriez explorer avec ce jeu de données ?

3. Types de variables :

- Utilisez les commandes comme `df.head()`, `df.info()`, et `df.describe()` pour observer la structure base des données ? Décrire le rôle de chacune de ces commandes ?
- Identifiez les types de variables et de données dans le jeu de données ?

4. Description statistique :

- Utilisez une méthode pour afficher un résumé statistique des colonnes numériques
- Quelle est la moyenne, l'écart type et les valeurs minimales et maximales des variables numériques ? Que pensez-vous ?

5. Valeurs manquantes :

- Utilisez une méthode pour vérifiez s'il y a des valeurs manquantes dans le jeu de données.
- Combien de valeurs manquantes y a-t-il dans chaque colonne en utilisant une méthode ?
- Proposer une méthode pour traiter ces valeurs manquantes ?

6. Valeurs aberrantes :

- Identifiez les valeurs qui semblent être des valeurs aberrantes (outliers) pour la colonne prix en utilisant des méthodes graphiques ?
- Proposez une manière de traiter ces valeurs aberrantes ?

7. Traitement des valeurs manquantes :

- Implémentez une méthode de traitement des valeurs manquantes. Expliquez pourquoi vous avez choisi cette méthode.

8. Transformation des données :

- Convertissez les variables catégorielles en format numérique si nécessaire (par exemple, en utilisant l'encodage).
- Expliquez pourquoi cette transformation est importante pour les étapes futures d'analyse.

9. Ajouter des markdown pour les cellules en utilisant les titres de niveau 1, des points [1-8] ?