

TP PANDAS énoncés

Préalable : réaliser la manip de <https://www.youtube.com/watch?v=zZkNOdBWgFQ>

Objectifs du TP

- Comprendre les bases de la bibliothèque Pandas.
- Savoir manipuler des DataFrame et des Series.
- Être capable d'effectuer des opérations de base telles que la lecture de fichiers, la sélection de données, le filtrage, et les modifications de DataFrame.
- Explorer des opérations plus avancées comme les jointures, les groupements, et les tris.

Contexte

Les données utilisées dans ce TP sont fictives et représentent les ventes d'une chaîne de magasins de vêtements. Le fichier de données (`donnees_ventes.csv`) contient les colonnes suivantes : Date, Magasin, Produit, Quantité vendue, et Prix unitaire.

Exercices

Exercice 1 : Préparation de l'environnement

Importez les bibliothèques nécessaires (`pandas` sous le nom `pd`).
Chargez le fichier `donnees_ventes.csv` dans un DataFrame `df`.

Exercice 2 : Exploration des données

Affichez les 5 premières lignes du DataFrame.
Affichez le nombre de lignes et de colonnes du DataFrame.
Affichez le type de chaque colonne.

Exercice 3 : Nettoyage des données

Vérifiez et traitez les valeurs manquantes.
Supprimez les éventuelles lignes dupliquées.

Exercice 4 : Manipulation de données

Ajoutez une nouvelle colonne `Revenu` qui est le produit de `Quantité vendue` par `Prix unitaire`.
Créez une colonne `Année` extraite de la colonne `Date`.

Exercice 5 : Sélection et filtrage

Sélectionnez les données des ventes qui ont eu lieu en 2022.
Filtrez les données pour afficher uniquement les ventes où plus de 100 articles ont été vendus.

Exercice 6 : Aggregations et statistiques

Calculez le total des revenus par magasin.
Trouvez le produit le plus vendu.

Exercice 7 : Tris et classements

Triez le DataFrame par date de vente en ordre décroissant.
Classez les magasins selon le revenu total en ordre décroissant.

Exercice 8 : Joins et fusions

Supposez que vous avez un autre DataFrame `df_promotions` avec les colonnes `Date`, `Magasin`, et `Promotion`. Effectuez une jointure avec `df` pour lier les données de promotions aux ventes.

Exercice 9 : Exportation de données

Exportez le DataFrame final en fichier CSV nommé `resultats_ventes.csv`.

Exercice 10 : Visualisation (Bonus)

Utilisez `matplotlib` ou `seaborn` pour visualiser les tendances des revenus au cours du temps.

QUESTIONS DE FOND ET COMPLÉMENTS ETUDE DE CAS

Exercice 11 : Poser puis résoudre 3 problèmes avec `scipy`. Exposé 2 résolutions à votre voisin / binôme Vous pouvez proposer une démonstration en salle de classe..

Exercice 12 : Commenter 3 codes sources utilisant `Pandas` et au moins une autre bibliothèque vu en classe.