Université Paris-Saclay
M1 ISD

Programmation en Python
2022-2023

Feuille d'exercices

Partie 5 : Pandas

Question 1 Definir dans une variable `schema_movies` une liste de paires `(clé, valeur)` dont les clés sont des noms de colonnes et les valeurs des chaînes de caractères représentant des types, pour le fichier `movies.csv`.

Question 2 Importer le module `pandas` et charger le fichier CSV `movies.csv` en indiquant que le séparateur de champs est `;` et que les types de colonnes sont définis dans `schema_movies`. Afficher la DataFrame obtenue avec la fonction `print`.

Question 3 [Statistiques] Calculer les valeurs suivantes, en utilisant les opérateurs de DataFrame et en sélectionnant judicieusement ces dernières.

- Nombre de lignes
- Année la plus grande
- Moyenne des durées

Question 4 [Filtrage] Pour chacun des critères suivants, calculer l'ensemble des lignes qui vérifient ces critères (et afficher ces lignes pour vérifier).

- Les films dont le rang est inférieur à 20
- Les films dont la durée est supérieure à la moyenne des durées
- Les films dont le titre contient `'Wars'`

Question 5 [Tris et sous-parties] Afficher les films triés par ordre croissants de rang et prendre les dix premières lignes. Afficher les films triés par ordre décroissant d'années et prendre les lignes 10 à 20.

Question 6 [Aggrégats] Compter le nombre de films par années. Compter la durée moyenne des films par année.

Partie 6 : Matplotlib

Question 7 Compter le nombre de films par années et afficher les résultats.

Le nombre d'années différentes étant trop grand pour avoir un graphique lisible, on va créer un résumé des données.

- Créer une copie de la DataFrame `movies`
- Sur la copie, mettre à jour la colonne `'YEAR'` en arrondissant chaque année à la décennie la plus proche. On peut par exemple faire la division entière par 10, puis remultiplier par 10.
- Effectuer un groupby par année et compter le nombre de films.
- Tracer un histogramme à partir de la DataFrame obtenue.

Question 8 Procéder comme ci-dessus pour afficher la moyenne des rangs des films par décennie.

Question 9 Afficher un graphique en nuage de points (`.scatter`) de la durée par rapport au rang du film. Peut on déduire une quelconque correlation ?