

Etude des films et séries disponibles sur Netflix

Contexte du projet

1. Lecture du fichier :
 - Dimension du dataframe
 - Combien y a-t-il de films dans ce jeu de données ?
 - Combien y a-t-il de séries ?
 - Générer le résumé statistique du dataframe
 - Compter les valeurs manquantes
2. Exploration des valeurs manquantes :
 - Sur la colonne des directeurs de production
 - Sur la colonne des acteurs
 - Suppression des lignes dupliquées (s'ils existent)
3. Combien de films/séries ont été produit par les Etats-Unis ? Par la France ?
4. Quel est le film avec la durée la plus longue sur Netflix ?
5. Etude des catégories avec le plus de contenu
6. Quels directeurs ont produit le plus de films/séries disponibles sur Netflix ?
7. Est-ce que Jan Suter travaille souvent avec les mêmes acteurs ?
8. Représenter les 10 pays qui ont produits le plus de contenus disponibles sur Netflix, avec le nombre de contenus par pays
9. Tracer un graphe à barres du nombre de films/séries par classement de contenu (rating)
10. Tracer un graphique qui montre l'évolution du nombre de films/séries produits par les 10 pays les plus producteurs de contenus sur Netflix, au fil des ans
11. Distribution de la durée des films disponibles sur Netflix
12. Tracer un graphique représentant le nombre de séries par modalité de nombre de saisons

Modalités pédagogiques

Pour réaliser ce projet, vous allez utiliser JupyterLab avec des librairies data science en python telles que Pandas, Matplotlib et SeaBorn.

Livrables

Un notebook comportant les analyses pré-exploratoires réalisées (non cleané, pour comprendre votre démarche). Interpréter vos résultats dans le notebook.