Projet Netflix - partie 2

12. Afficher les directeurs qui ont produit le plus de films/séries disponibles sur Netflix

Action : adapter le code précédent pour la colonne director plutôt que la colonne listed_in, afin d'afficher les dix directeurs de production les plus présents dans le jeu de données Netflix.
Indices :

- Remplacer le nom de colonne listed_in par director dans le code de question 11.

13. Voir si Jan Suter travaille souvent avec les mêmes acteurs

Action : afficher les cinq acteurs les plus présents dans les contenus produits par Jan Suter et en déduire si ce directeur travaille préférentiellement avec certains acteurs ou non.
Indices :

- Garder uniquement les lignes du dataframe où les valeurs de la colonne director ne sont pas manquantes, grâce à la méthode notna().

- Utiliser la méthode str.contains() pour sélectionner uniquement les lignes dont la colonne director contient la chaîne de caractères "Jan Suter" et ranger ce nouveau dataframe dans une variable nommée, par exemple, donnees_jan_suter.

- Adapter le code précédent, utilisé pour les colonnes listed_in et director, afin d'afficher les cinq acteurs les plus présents dans le dataframe donnees_jan_suter.

14. Représenter les dix pays qui ont produit le plus de contenus disponibles sur Netflix, avec le nombre de contenus par pays

Action : écrire le code pour sélectionner les dix pays avec le plus de contenus, créer un sous-dataframe contenant uniquement les données de ces pays, puis tracer un countplot sur ce dataframe et la variable country.
Indices :

- Sélectionner les dix pays ayant produit le plus de contenus grâce à la méthode value_counts() et la méthode head() et stocker cette série dans une variable.

- Créer un dataframe contenant uniquement les lignes pour ces dix pays, c'est-à-dire les lignes pour lesquelles la valeur dans la colonne country est comprise dans les index de la série créée précédemment. Pour cela, utiliser l'indexation booléenne, la méthode .isin() et l'attribut index.

- Tracer un countplot grâce à la fonction countplot() de Seaborn en lui donnant le nouveau dataframe et la variable souhaitée : country.

15. Tracer un graphe à barres du nombre de films/séries par classement de contenu (rating)

Action : tracer le graphe à barres en utilisant le dataframe d'origine et la colonne d'intérêt représentant le classement de contenu.
Indice :

- La fonction pour tracer un graphe à barres est countplot().

- La colonne utile est la colonne rating.

16. Afficher l'évolution du nombre de films/séries disponibles sur Netflix au cours du temps

a. Notions supplémentaires sur les dates

Il est possible de formater une colonne date d'un dataframe grâce à la fonction to_datetime() de Pandas, qui permet de la passer dans le type datetime64.

Il existe des attributs pour les données de type datetime64 qui permettent d'accéder à des valeurs spécifiques, comme le jour dt.day, le mois dt.month ou encore l'année dt.year.

Syntaxe pour l'utilisation de ces attributs :

```
dataframe["colonne"].dt.day
dataframe["colonne"].dt.month
dataframe["colonne"].dt.year
```

b. Énoncé

Action : écrire le code pour formater correctement la colonne date_added et créer une nouvelle colonne dans le dataframe contenant uniquement l'année d'ajout du contenu sur Netflix. Grouper le dataframe sur les variables year_added et type, afin de tracer l'évolution du nombre de contenus ajoutés par an grâce à la fonction pointplot() de Seaborn. Séparer les courbes selon le type du contenu, Movie ou TV Show.

Indices :

- La méthode to_datetime() permet de formater la colonne date_added, date d'ajout du contenu sur Netflix.

- À partir de la colonne date_added correctement formatée, récupérer l'année d'ajout du contenu sur Netflix avec l'attribut dt.year et créer une nouvelle colonne dans le dataframe, nommée par exemple year_added.

- Grouper le dataframe sur les variables souhaitées grâce à la méthode groupby() de Pandas, déterminer le nombre d'occurrences par groupe grâce à la méthode size() puis renommer la colonne de comptage en Count. Un exemple est disponible dans la section sur la fonction pointplot() du chapitre Seaborn.

- Utiliser la fonction pointplot() de Seaborn avec en x la variable year_added, en y la variable Count et comme option hue la variable type, afin d'avoir une courbe par type de contenu.

17. Afficher la distribution de la durée des films disponibles sur Netflix

Action : écrire le code pour sélectionner uniquement les contenus correspondant au type Movie et transformer la durée des films en nombre entier, comme dans la section Afficher le film avec la durée la plus longue sur Netflix de ce chapitre. Enfin, tracer un histogramme sur la série contenant les données de durée des films.
Indices :

- Créer une série contenant les durées de films en nombres entiers.

- Utiliser la fonction distplot() de Seaborn pour tracer la distribution de ces durées.

18. Tracer un graphique représentant le nombre de séries par modalité de nombre de saisons

Action : créer un dataframe des contenus de type TV Show uniquement puis tracer un countplot() pour représenter le nombre de séries pour chaque modalité du nombre de saisons.
Indices :

- Utiliser la sélection booléenne pour créer le dataframe des séries.

- La variable utile en x est la variable duration, contenant l'information du nombre de saisons pour chaque série.