

Olivier Seguin - SEGO27120400

Elliot Shepherd - SHEE15060600

Olivier Robin - ROBO10050700

Outils de programmation pour la science des données
8PRO408

Analyse des contenus Netflix

Travail présenté à
Habba's Ngodjou Doukaga

Département d'informatique et de mathématique
Université du Québec à Chicoutimi
10 décembre 2025

Résumé des observations principales

1. Valeurs manquantes

Le dataset comporte plusieurs colonnes incomplètes :

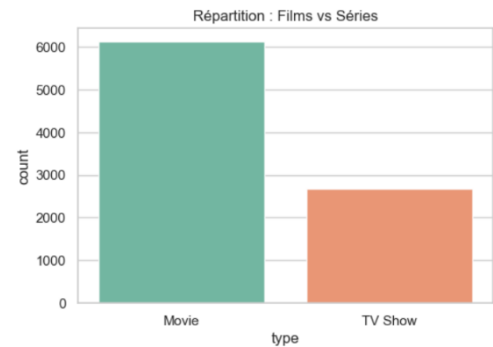
- *directors* : Une grande proportion de titres n'ont aucun réalisateur indiqué
- *cast* : de nombreux titres n'ont pas de distribution renseignée.
- *country* : plusieurs titres n'ont aucun pays de production spécifié.
- *date_added* : quelques valeurs sont manquantes.

2. Films vs Séries

Le dataset est composé de deux types de contenu :

- 69,62 % de films
- 30,38 % de séries

Les films sont beaucoup plus nombreux, ce qui influence plusieurs analyses (genres, durées, etc.).

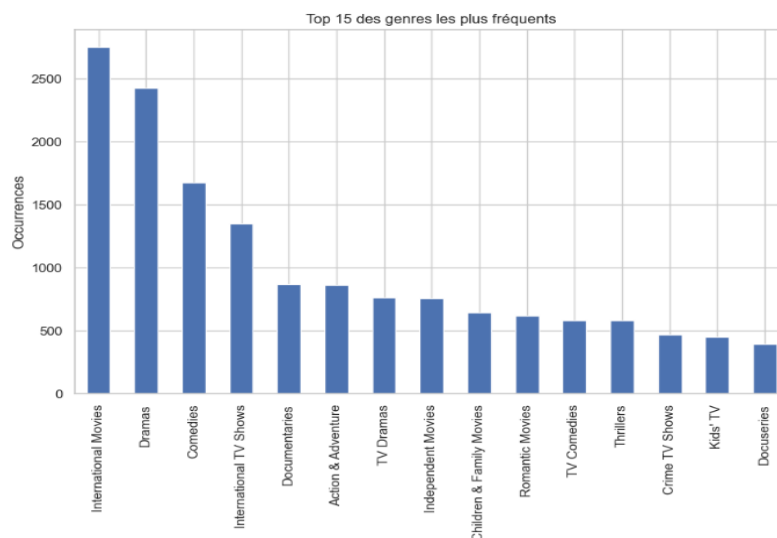


3. Genres

Plusieurs genres sont plus communs que d'autres :

- *Films et séries internationales*
- *Drames*
- *Comédies*

Ces catégories sont très larges et représentent beaucoup de titres, ce qui explique qu'ils sont plus fréquents. De plus, la majorité des titres possèdent plusieurs genres, ce qui augmente le nombre d'occurrences de plusieurs genres.



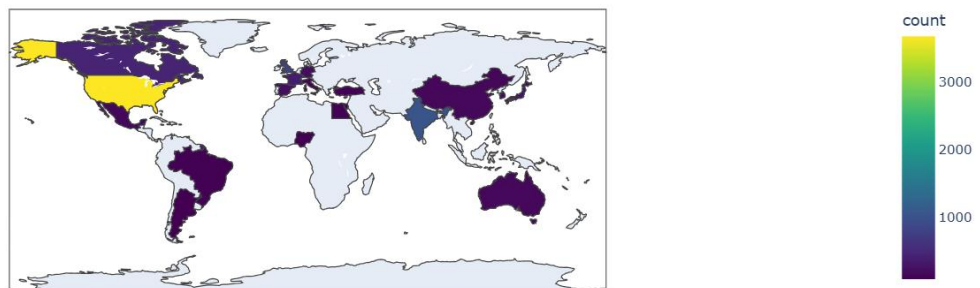
4. Répartition géographique

La répartition des pays producteurs est très déséquilibrée :

- Les États-Unis ont produit la majorité des titres (41,89%).
- Ils produisent environ 3,5 fois plus de titres que le deuxième pays (l'Inde).
- Cette concentration rend certaines visualisations (comme les heatmaps) plus difficile à interpréter, car il y a un seul pays dont la couleur contraste des autres.

*Nous n'avons pas fait une heatmap de tous les pays producteurs, car les États-Unis écrasaient les autres pays et nous avions un problème avec les couleurs.

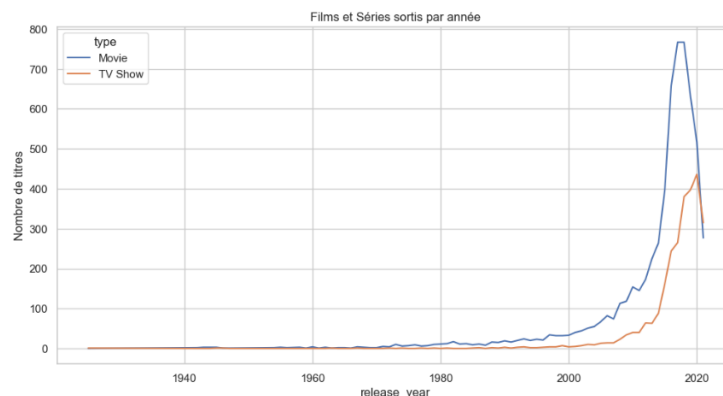
Top 20 des pays producteurs (Carte interactive)



Tendances remarquables

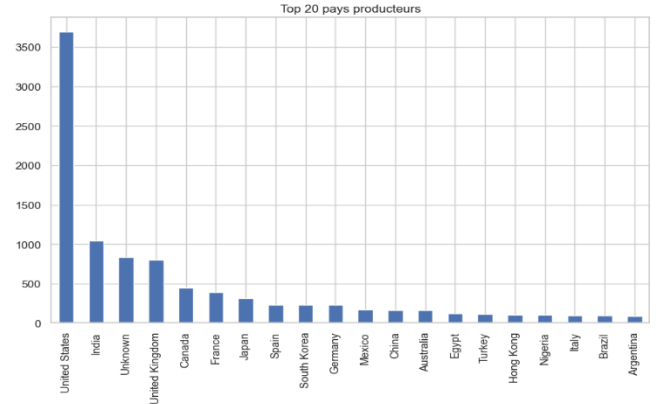
1. Évolution temporelle des sorties (release_year)

- Le nombre de titres sortis augmente beaucoup au fil des décennies, jusqu'à atteindre un maximum autour de 2018–2020.
- Autour de 2020–2021, il est possible de constater une diminution de sortie de titres pour potentiellement deux raisons :
 - i. la pandémie de COVID-19, qui a ralenti plusieurs productions,
 - ii. le fait que le dataset ne couvre pas toute l'année 2021 (dernière mise à jour : 25 septembre 2021).



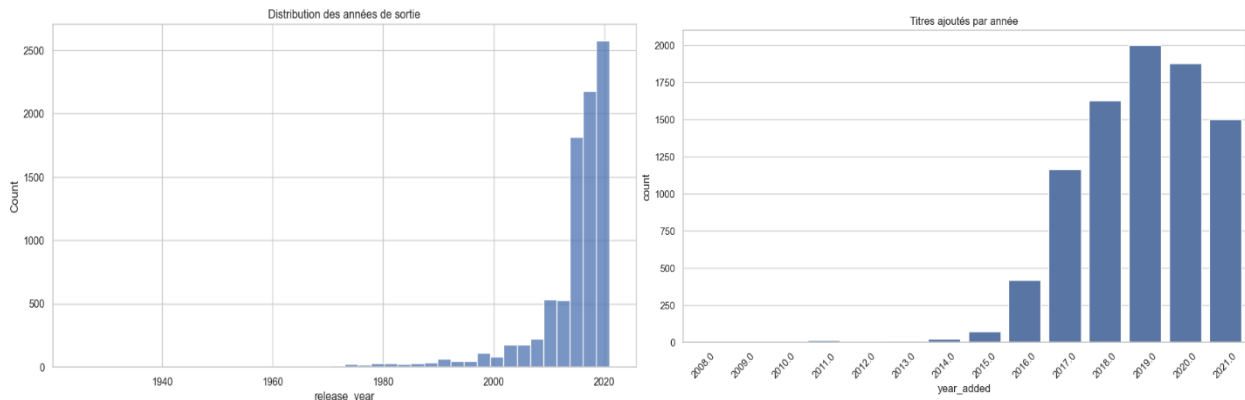
2. Pays dominants

- Les 20 pays qui produisent le plus de titres sont principalement des pays développés.
- Il est possible de constater que certains pays produisent plus que les autres comme les États-Unis, l'Inde et le Royaume-Uni.



3. Titres ajoutés sur Netflix (date_added)

- La majorité des titres ajoutés sur Netflix sont des films et séries sorties récemment.
- Dans la majorité des cas, la date d'ajout sur Netflix est peu après la date de sortie du titre.



4. Durée des films

Bien que la durée semble homogène dans certains graphiques, les valeurs sont relativement dispersées :

- La grande majorité se situe entre 47 minutes et 154 minutes ce qui est un grand écart.
- Quelques valeurs extrêmes augmentent aussi la variabilité. (Ex : 3min et 312min)

