

Adèle Dalle

10/12/2025

Outils Programmation pour la science des données

Analyse exploratoire du catalogue Netflix (Movies & TV Shows)

Dans le cadre de ce projet, j'ai analysé le jeu de données `netflix_titles` provenant de la plateforme Kaggle afin de pouvoir produire des visualisations et représentations graphiques sur le catalogue Netflix.

J'ai réalisé ce projet en plusieurs étapes. J'ai tout d'abord chargé et nettoyé les données. J'ai ensuite analysé les contenus et effectué une analyse temporelle sur certaines variables avant de produire des visualisations interactives et une petite application streamlit.

Lors du nettoyage des données, j'ai supprimé les lignes vides ainsi que les doublons. J'ai ensuite converti les colonnes `date_added` et `release_year` en datetime.

Après ce nettoyage, je me suis concentrée sur l'analyse des contenus. On observe que les films sont le type de contenu le plus présent sur la plateforme Netflix. De plus, on observe un ajout exponentiel de contenu à partir des années 2010, ce qui correspond à la période d'introduction du streaming en ligne par la plateforme, avec un pic en 2017 suivi d'une légère baisse d'ajout de contenu dans les années suivantes. Netflix propose 15 genres principaux de contenu entre des films internationaux, des documentaires et des comédies musicales. Les genres les plus fréquent sur la plateforme sont les films internationaux (2369), les drames (2293) et les comédies (1553). Les contenus disponibles sur la plateforme sont produits par de nombreux pays. Les pays principaux sont les Etats-Unis (2485), l'Inde (940) et le Royaume Unis (484). De nombreux acteurs et actrices sont donc amenés à travailler dans les contenus présents sur la plateforme. Les acteurs et actrices que l'on retrouve le plus sur la plateforme sont Anupam Kher (42 films), Shah Rukh Khan (34 films) et Naseeruddin Shah (31 films). De même, les réalisateurs et réalisatrices les plus présents sont Jan Suter (21 films), Raúl Campos (19 films), Marcus Raboy (15 films) et Jay Kars (15 films).

J'ai ensuite effectué une analyse temporelle sur les années de sorties ainsi que les dates d'ajout des contenus. On observe une sortie exponentielle de contenu à partir des années 2010, ce qui correspond à la période d'introduction du streaming en ligne par la plateforme, avec un pic en

2017 suivi d'une légère baisse de sortie de contenu dans les années suivantes. On observe que de nombreux ajouts de contenus ont été fait à partir de 2017 avec un pic en 2019 suivi d'une légère recrudescence d'ajout dans les années suivantes. Les mois les plus propices à l'ajout de contenu sur la plateforme sont janvier, octobre et décembre. Les mois durant lesquels le moins d'ajout de contenu est effectué sont février, mai, juin et septembre.

Puis j'ai produit des visualisations tels qu'un histogramme, un countplot, des boxplots et des visualisations interactives avec Plotly. On observe que la majorité des contenus proposés sont compris entre 80 et 120 minutes (en excluant les séries). Pour les séries, le temps le plus proposé est 1 saison. Concernant la classification des contenus, la classification la plus présente est TV-MA (contenu mature), suivi d'un contenu déconseillé au moins de 14 ans (TV-14). D'après la boîte à moustaches de la durée des films en minutes, on observe que la majorité des films durent entre 80 et 120 minutes. Il y a quelques films dépassant les 200 minutes mais ils restent présents en minorité. Concernant les séries, la majorité des séries présentes sur la plateforme Netflix comptent une seule saison. Quelques séries comptent plus de 4 saisons et une série compte plus de 14 saisons mais elles restent minimales. Nous pouvons observer la répartition des contenus par genre. Les films internationaux représentent 23% de l'ensemble des contenus alors que les films indépendants ne représentent que 7%.

L'analyse du jeu de données met en évidence la manière dont Netflix a construit et fait évoluer son catalogue. Après le nettoyage des données, on observe que les films dominent largement la plateforme, avec une durée moyenne comprise entre 80 et 120 minutes et une classification visant davantage un public adulte (TV-MA). Le nombre de films et de séries dans le catalogue s'accélère dès 2010, ce qui correspond à l'essor du streaming, avec un pic en 2017 pour les sorties et en 2019 pour les ajouts, suivi d'une stabilisation. La diversité des genres et des pays producteurs montre la volonté de Netflix d'être perçu comme une plateforme internationale. Elle est dominée par les États-Unis, l'Inde et le Royaume-Uni, mais reste ouverte à de nombreux autres pays. La présence récurrente de certains acteurs/actrices et réalisateurs/réalisatrices souligne également l'importance de figures clés dans la constitution du catalogue. Netflix s'est donc imposé grâce à une offre large, variée et mondiale, construite sur une croissance rapide puis stabilisée, et orientée vers un public adulte.