

Université du Québec à Chicoutimi (UQAC)

Département d'informatique et de mathématique

8PRO408 - Outils de programmation pour la science des données

Rapport d'Analyse Exploratoire : Catalogue Netflix (Movies & TV Shows)

Présenté par :
Thibault MARTIN
Oscar NEVEUX

Code source et analyses :
<https://github.com/LaPoulePunto/netflix-eda-analyse>
Site Streamlit :
<https://lapoulepunto-netflix-eda-analyse-app-obtr78.streamlit.app/>

Date :
7 décembre 2025

1 Introduction

Ce rapport présente les résultats de l'analyse exploratoire des données (EDA) effectuée sur le catalogue de la plateforme de streaming Netflix. L'objectif de cette étude est de comprendre la composition du catalogue, d'identifier les stratégies d'acquisition de contenu au fil du temps et de mettre en lumière les caractéristiques géographiques des productions.

Le jeu de données, arrêté à l'année 2021, contient **8807 entrées** et 12 attributs descriptifs. Ces attributs incluent le type de contenu (film ou série), le titre, le réalisateur, le casting, le pays de production, la date d'ajout sur la plateforme, l'année de sortie réelle, la classification d'âge, la durée et les genres associés.

2 Qualité des Données et Prétraitement

Avant de procéder à l'analyse, un audit de la qualité des données a été réalisé. Le jeu de données est globalement propre mais présente certaines lacunes spécifiques :

- **Valeurs manquantes** : La colonne `director` est la plus affectée avec environ 30 % de valeurs nulles, suivie par `country` (9 %) et `cast` (9 %). Pour l'analyse géographique, les valeurs manquantes ont été étiquetées comme "Unknown" afin de ne pas perdre d'informations sur les autres dimensions.
- **Conversion temporelle** : La colonne `date_added` a été convertie en format `datetime` pour permettre l'extraction de l'année et du mois d'ajout.
- **Intégrité** : Aucune ligne dupliquée n'a été détectée.

3 Analyse de la Répartition des Contenus

3.1 Déséquilibre Films vs Séries

L'analyse de la variable `type` révèle une prédominance historique des longs métrages sur la plateforme.

TABLE 1 – Répartition des contenus par type

Type	Nombre de titres	Pourcentage
Movies (Films)	6128	69.6 %
TV Shows (Séries)	2676	30.4 %

Bien que Netflix soit souvent médiatisé pour ses séries originales, le volume de films reste supérieur, servant probablement de base de contenu "fonds de catalogue" pour retenir les abonnés entre deux saisons de séries populaires.

4 Dynamique Temporelle

4.1 Année de Sortie vs Date d'Ajout

L'analyse montre une accélération massive de l'enrichissement du catalogue à partir de 2015-2016. Le pic d'ajouts se situe entre **2018 et 2020**. On observe un léger ralentissement en 2021, potentiellement attribuable aux effets de la pandémie de COVID-19 sur les chaînes de production ou à une saturation du marché.

Concernant l'âge des contenus, la stratégie de Netflix est résolument tournée vers la nouveauté. La médiane des années de sortie se situe autour de 2017. Bien que le catalogue contienne des classiques (le plus ancien datant de 1925), la grande majorité des titres a moins de 10 ans d'ancienneté.

5 Analyse Géographique et Genres

5.1 Contenu Américain et Diversification

Les États-Unis dominent largement la production, étant le pays d'origine principal pour plus de 2800 titres. Cependant, la stratégie d'internationalisation de Netflix est visible à travers la forte présence de l'Inde (Bollywood) et du Royaume-Uni.

Le tableau ci-dessous présente les 5 principaux pays producteurs identifiés dans le jeu de données :

TABLE 2 – Top 5 des pays producteurs (hors inconnus)

Pays	Nombre de titres
États-Unis	3687
Inde	1046
Royaume-Uni	806
Japon	445
Corée du Sud	231

La présence notable du Japon et de la Corée du Sud confirme l'attrait mondial pour les productions asiatiques comme les animés, K-dramas.

5.2 Genres et Public Cible

L'analyse des genres (variable `listed_in`) montre que Netflix vise un public large mais segmenté. Les catégories les plus fréquentes sont :

1. **International Movies (2752 titres)** : Cette catégorie confirme la volonté d'exporter des films locaux vers une audience globale.
2. **Dramas (2427 titres)** : Le genre dramatique reste la valeur sûre du catalogue.
3. **Comedies (1674 titres)** : Pour le divertissement grand public.

En croisant les données de classification d'âge (`rating`), on constate une prédominance des contenus classés **TV-MA** (Mature Audience) et **TV-14**. Cela indique que le catalogue cible principalement un public adulte et jeune adulte, plutôt que les enfants, bien que la section "Kids" existe.

6 Conclusion

Cette analyse exploratoire permet de dresser le profil type du contenu Netflix actuel : une œuvre récente (post-2016), majoritairement américaine ou indienne, et orientée vers un public adulte (Drames/Comédies).

L'évolution temporelle suggère que la phase d'acquisition massive de volume (2016-2020) pourrait laisser place à une stratégie de consolidation ou de production originale plus ciblée. La distinction nette entre le volume de films et les séries reste une caractéristique structurelle de l'offre.

Il serait pertinent pour un autre projet de croiser ces métriques descriptives avec des données de popularité ou des notes critiques (IMDb/Rotten Tomatoes) pour évaluer si la quantité de contenu corrèle avec la qualité perçue par les utilisateurs.