

Analyse Exploratoire du Catalogue Netflix (EDA)

1. Introduction

Ce rapport présente une analyse exploratoire du dataset « Netflix Movies and TV Shows », dans le cadre du mini-projet du cours 8PRO408 – Outils de Programmation Python.

L'objectif principal est de comprendre la composition du catalogue Netflix à travers des statistiques descriptives, du nettoyage de données et des visualisations (Matplotlib, Seaborn et Plotly).

2. Méthodologie

Le traitement des données a été réalisé en Python avec les bibliothèques Pandas, NumPy, Matplotlib, Seaborn et Plotly. Les étapes suivies sont :

- Importation du dataset brut fourni au format CSV.
- Nettoyage : suppression des doublons, harmonisation des noms de colonnes, traitement des valeurs manquantes, extraction du genre principal et du pays principal.
- Transformation : conversion des dates, extraction de l'année d'ajout, création de nouvelles colonnes.
- Visualisation : exploration interactive et statique des films, séries, genres, pays et tendances temporelles.

3. Résultats Principaux

L'analyse met en évidence plusieurs tendances :

- Les films sont légèrement plus nombreux que les séries dans le catalogue Netflix.
- Les genres les plus représentés incluent Drama, Comedy, Documentary et International Movies.
- Les pays les plus présents dans le catalogue sont les États-Unis, l'Inde et le Royaume-Uni.
- Le nombre de titres ajoutés augmente fortement après 2015, reflétant l'expansion internationale de la plateforme.
- On observe un grand nombre de valeurs manquantes dans les colonnes director, cast et country.

4. Conclusion

Cette analyse exploratoire montre que Netflix propose un catalogue diversifié dominé par les films et des genres populaires tels que le drame et la comédie. Le dataset révèle également une forte croissance du nombre de titres après 2015. Malgré cela, certaines limites restent présentes, notamment l'absence d'informations sur la popularité réelle des

titres, des données partielles pour certaines colonnes, et la nature statique du dataset, qui ne reflète pas les évolutions continues du catalogue Netflix.

5. Limites

- Le dataset ne contient aucune donnée sur la popularité ou l'engagement des utilisateurs.
- Les dates ajoutées au catalogue ne sont pas toujours précises.
- Certaines colonnes textuelles nécessiteraient un nettoyage plus avancé.
- Le dataset représente un instant figé et ne permet pas d'analyser l'évolution du catalogue dans le temps.