

# Rapport d'Analyse du Notebook : Exploration du Catalogue Netflix

## 1. Introduction

Ce rapport présente une synthèse du notebook consacré à l'analyse exploratoire du dataset « Netflix Movies & TV Shows », contenant 8 807 titres. L'objectif du notebook est de mieux comprendre la composition du catalogue Netflix à travers une exploration des données, des visualisations et des statistiques descriptives.

## 2. Chargement et exploration initiale du dataset

Le notebook débute par l'importation des bibliothèques nécessaires, notamment pandas, numpy et matplotlib. Le dataset est ensuite chargé depuis un fichier CSV. Une première inspection permet d'obtenir le nombre total de lignes et de colonnes, un aperçu des premières entrées ainsi que des informations générales sur les types de données présents dans chaque variable.

L'analyse met en évidence la structure globale du dataset, comprenant des colonnes telles que le title, les directeurs, le casting, c'est-à-dire les acteurs du film, les pays d'origine, la date d'ajout sur Netflix, etc. Une section est également consacrée à la vérification des valeurs manquantes afin de déterminer les champs nécessitant un prétraitement. On a alors décidé de modifier les valeurs manquantes de nos colonnes d'intérêt par « Non spécifié » et « Non classifié ».

## 3. Analyse descriptive

Des statistiques descriptives sont générées pour les variables pertinentes. Celles-ci incluent notamment la distribution des années de sortie des contenus, la répartition des catégories « Movies » et « TV Shows », ainsi que les pays les plus représentés dans le catalogue.

Notre notebook procède également à l'analyse de la colonne “listed\_in” afin d'examiner les genres les plus populaires. Cette étape met en lumière les préférences en termes de catégories, confirmées plus tard par les visualisations.

## **4. Visualisations**

Plusieurs graphiques sont produits dans le notebook :

- Répartition des types de contenus (films vs séries).
- Histogramme des années de sortie, montrant une forte concentration de titres récents.
- Classement des pays les plus représentés.
- Genres les plus fréquents dans le catalogue.
- Analyse du rating (ex : TV-MA, PG-13, etc.).

Ces visualisations permettent de comprendre rapidement les grandes tendances du catalogue Netflix, notamment l'augmentation du volume de contenus originaux au fil des années et la prédominance des productions américaines.

## **5. Conclusions**

L'analyse montre que le catalogue Netflix est majoritairement composé de films, avec une proportion significative de séries originales. Les contenus sont récents pour la plupart, avec une forte représentation du marché américain. Les genres les plus répandus incluent les documentaires, les drames et les comédies.

Ce notebook constitue une base solide pour une étude plus approfondie, incluant des modèles de recommandation, des analyses temporelles ou une exploration textuelle des descriptions.