



2025-12-10

Analyse du jeu de donnée - Netflix

8PRO408 - Outils programmation pour la science des données



Jérémie Grenon, Mathieu Dionne et Alexandre Perron
UNIVERSITÉ DU QUÉBEC À CHICOUTIMI

Table des matières

Description	2
Technologies utilisées	2
Travail effectué	2
Exploration du jeu de données	2
Analyse des contenus.....	2
Analyse temporelle	3
Visualisations	3
Synthèse.....	3
Contenus récents	3
Pays dominants	3
Duré des films	3
Genres populaires	3
Production.....	3
Conclusion.....	3

Description

Ce projet consiste à l'exploration et la visualisation du jeu de donnée du catalogue Netflix 2021. Dans ce travail, on explore la structure et la qualité du jeu de données. On étudie le type de contenu proposé, la distribution des genres, pays et années de sortie ainsi que les tendances observées. On propose également une visualisation intéressante ainsi qu'une synthèse globale du projet.

Technologies utilisées

Pour effectuer la tâche, nous avons utilisé VS Code avec l'extension Jupyter Notebook afin de rédiger l'analyse et le tri du jeu de donnée. Voici également quelques bibliothèques utilisées :

- Pandas (pour la manipulation de données)
- Matplotlib (Pour la visualisation)
- Seaborn (Pour la visualisation)
- Streamlit (Pour le déploiement de l'application)
- Streamlit_jupyter (Pour la liaison Jupyter Notebook -> streamlit)

Afin d'être facilement visualisée par n'importe quel utilisateur, l'application est déployée sur Streamlit, une plateforme permettant d'héberger une application web gratuitement à des fins personnels, éducatives ou autres. Voici la [page web du projet](#) ainsi que le [dépôt Github](#).

Travail effectué

Le jeu de donnée est tout d'abord lu sous forme de fichier .csv et converti sous forme de « dataframe » (pandas) afin de pouvoir l'utiliser facilement dans toute l'application.

Exploration du jeu de données

- Liste des différentes colonnes
- Affichage des valeurs manquantes par colonne sous forme de tableau
- Affichage du nombre de duplcats
- Affichage des différents types de données sous forme de tableau
- Nettoyage des données

Analyse des contenus

- Affichage du nombre total de films et de séries ainsi que leur pourcentage
- Affichage du nombre de films et séries par année de sortie
- Affichage interactif du nombre de films et séries par catégorie sous forme de tableau
- Affichage interactif du nombre de films et séries par pays sous forme de tableau
- Affichage interactif du nombre d'apparition par acteur sous forme de tableau et graphique
- Affichage interactif du nombre de réalisation par réalisateur sous forme de tableau et graphique

Analyse temporelle

- Affichage interactif de la distribution des contenus par année de sorties sous forme de graphique
 - Nombre de titres par année de sortie
 - Répartition des films et séries par année de sortie
- Affichage des titres ajoutés sur Netflix par années sous forme de graphique
- Zoom sur les ajouts récents
 - Affichage interactif du nombre de films et série par année depuis l'année sélectionnée sous forme de graphique
- Affichage du nombre de films vs séries sous forme de graphique
- Distribution de la durée des films en minutes sous forme de « Boxplot »
- Affichage des 5 catégories les plus populaires sous forme de diagramme circulaire

Visualisations

Différents graphiques sont présents un peu partout dans l'application.

Synthèse

Contenus récents

La majorité des titres datent des années récentes, surtout après 2010. Les années 2016 à 2019 sont celles où Netflix a ajouté le plus de nouveaux contenus.

Pays dominants

Les États-Unis dominent largement le catalogue, suivis par l'Inde. On observe également une croissance notable des productions internationales, signe que Netflix élargit son offre mondiale.

Durée des films

On observe que la plupart des films ont une durée d'environ 90 minutes, ce qui correspond au format le plus répandu dans l'industrie. La distribution est assez concentrée autour de cette valeur, avec quelques « outliers » représentant des films très courts ou beaucoup plus longs.

Genres populaires

Les catégories les plus représentées sont « *International Movies* », « *Dramas* », « *Comedies* » et les contenus pour enfants. Ce sont les genres que Netflix semble pousser le plus.

Production

Le nombre de productions a explosé à partir de 2015. On voit clairement que le catalogue atteint un sommet autour de 2019. Rien n'est comparable avec les années 2000 et les années précédentes, où le nombre de titres ajoutés était beaucoup plus faible. Cela montre bien l'expansion massive de Netflix durant cette période.

Conclusion

Netflix propose un catalogue majoritairement récent, dominé par les productions américaines.