

# Projet: Analyse Exploratoire des Données (Exploratory Data Analysis)

L'analyse exploratoire des données (EDA) est une méthode pour explorer un ensemble de données et découvrir des tendances, des relations et des modèles qui pourraient être cachés à première vue. Cela implique de regarder les données sous différents angles et de résumer les résultats de manière visuelle pour mieux comprendre les caractéristiques des données.

Le projet consiste à étudier un des trois datasets : **Netflix.csv**, **Amazon.csv**, **Disney.csv**

Disponibles sur : <https://github.com/EMSIMa/ADD3IIR/tree/main/Projet/Datasets> (consulter le document sur google classroom pour connaître quel dataset vous a été assigné).

**Vous devez rendre un fichier .ipynb** contenant du code et du texte pour répondre aux questions suivantes:

- Top 5 des meilleures catégories.
- Top 5 des réalisateurs.
- Les dix meilleurs acteurs.
- Les cinq meilleures séries télévisées avec le plus grand nombre de saisons.
- Identifier des contenus similaires en faisant correspondre des caractéristiques textuelles
- Netflix/Amazon/Disney se concentre-t-il davantage sur les séries télévisées que sur les films ces dernières années ?

Pour répondre à ces questions vous devez utiliser les outils que nous verrons en cours (pandas et matplotlib/seaborn)

Votre Jupyter Notebook doit contenir du code qui réponds à ces questions ainsi que du code pour des visualisations de ces données. Donc vous devez écrire du code et expliquer chaque cellule de code avec une cellule de texte, vous devez écrire du code pour toutes les étapes qui vous ont amenées à répondre aux questions ci-dessus.

Votre fichier ipynb doit également contenir des visualisations des données étudiée.

Note : Vous êtes libres d'explorer les données au-delà des questions ci-dessus, cela vous confèrera sans doute un bonus.

## Rendu du projet et règles de soumissions :

- Chaque équipe doit rendre un fichier *.ipynb*.
- Consultez Google Classroom pour voir le numéro de votre Equipe et le dataset assigné
- Le fichier du rendu final doit être nommé de la manière suivante :  
*Groupe\_NumeroEquipe.ipynb* Exemple : **G3\_04.ipynb**

- La première cellule du fichier .ipynb doit contenir les noms des membres de l'équipe ainsi que leur groupe
- Ce fichier doit être soumis **avant la date limite** via ce [formulaire](#)
- La **Date Limite** pour rendre le projet est le **Jeudi 16 Mai 2024 à 23:59:59**.
- Au-delà de la date limite, toutes **24 heures de retard** engendrera un **malus de 2 points** pour l'équipe.

## Ressources :

Pour vous donner un peu une idée sur l'analyse exploratoire des données, je mets à votre disposition les ressources suivantes :

- Vidéo en Darija : <https://youtu.be/7pVSitAcGYY> (les 46 Premières minutes)
- Vidéo en Français : <https://youtu.be/KDzUozgLyCl>
- Vidéo en Anglais : <https://youtu.be/xi0vhXFPegw>