

**Département d'informatique et de mathématique**  
**8PRO408 - Outils de programmation pour la science des données**  
**Chargé de Cours : HN Doukaga, hndoukag@uqac.ca**

**Mini-Projet : Analyse des contenus Netflix**

**Analyse exploratoire du catalogue Netflix (Movies & TV Shows)**

**Description générale**

Vous utiliserez un jeu de données réel provenant de Netflix (8 807 titres, 12 colonnes) afin de réaliser une **analyse exploratoire structurée** du catalogue : films, séries, pays d'origine, genres, dates d'ajout, casting, etc.

L'objectif du mini-projet est de comprendre la composition du catalogue Netflix, d'explorer sa diversité et d'identifier des tendances clés (types de contenus, répartition géographique, évolution temporelle, etc.).

**Objectifs du projet**

- Explorer la structure et la qualité du dataset.
- Analyser les types de contenus proposés (films vs séries).
- Étudier la distribution des genres, pays, années de sortie.
- Identifier des tendances temporelles (contenus récents, ajouts par année, etc.).
- Produire des visualisations pertinentes à l'aide de Pandas, Seaborn, Matplotlib et Plotly.
- Résumer les observations sous forme d'un court rapport analytique.

**Travail demandé**

1. **Exploration du dataset**
  - Aperçu général : colonnes, valeurs manquantes, duplicats, types de données.
  - Vérification et nettoyage minimal (si nécessaire).
2. **Analyse des contenus**
  - Films vs séries : proportions, tendances par année.
  - Genres principaux (listed\_in) : regroupement, fréquences.
  - Répartition géographique (country).
  - Casting & réalisateurs : analyse simple (comptages, noms fréquents).
3. **Analyse temporelle**
  - Distribution par année de sortie (release\_year).
  - Analyse de la colonne date\_added.
4. **Visualisations**
  - Représentations obligatoires :
    - histogrammes, countplots, boxplots (Seaborn / Matplotlib)
    - visualisations interactives (Plotly)
  - Choix libres : nuages de mots, diagrammes circulaires, timelines, etc.
5. **Synthèse**
  - Résumé des observations principales.

- Tendances remarquables (contenus récents, pays dominants, genres populaires).

## **Livrables**

À déposer sur GitHub :

- Un **notebook Jupyter** complet avec analyses et graphiques.
- Un **rapport PDF court** (1 à 2 pages).
- Une **mini application Streamlit** (visualisations interactives).
- Un fichier README.md expliquant comment exécuter le projet.

## **Modalités**

- Travail individuel ou en équipes de 2 à 4.
- Date limite : **10 décembre 2025 – 23h59**.
- Le dépôt GitHub doit être accessible.

## **Évaluation (100 points)**

- Qualité de l'analyse exploratoire /**25**
- Variété et pertinence des visualisations /**15**
- Clarté du notebook /**25**
- Qualité du rapport /**15**
- Fonctionnalité de l'application Streamlit /**10**
- Organisation du dépôt GitHub /**10**

## **Pénalités**

- Mauvais français : -20 %
- Retard : -10 %, puis -5 % par jour (max 7 jours)
- Non-respect des consignes GitHub : -5 à -15 %

**Liens du jeu de données : [Kaggle](#)**