

Projet de la Session — Dashboard Analyse de Films IMDb

Objectif

Construire un dashboard fiable, lisible et exploitable permettant d'analyser le **Top 250 des meilleurs films IMDb** à partir de données réelles collectées en ligne.

Le dashboard doit permettre de répondre à des questions concrètes :

- Quels genres dominent le Top 250 ?
- Quels réalisateurs sont les plus représentés ?
- Comment évoluent les notes dans le temps (par année / décennie) ?
- Quels films combinent **meilleure note** et **fort volume de votes** ?
- Observe-t-on une tendance de “qualité” au fil des années ?

Source des données

Les données doivent être collectées depuis :

👉 <https://www.imdb.com/chart/top/>
(Page du Top 250 des meilleurs films selon IMDb)

Mission

Réaliser un dashboard qui :

- repose sur des **données réelles** scrapées depuis IMDb,
- propose une **lecture structurée** du Top 250,
- permet d'explorer les films via des filtres,
- transforme les données brutes en **insights actionnables**.

Niveau de précision attendu

Le dashboard doit permettre d'analyser précisément :

- **Genres** : répartition, top genres, co-occurrences
- **Réalisateurs** : top réalisateurs, fréquence
- **Notes** : distribution, tendances
- **Votes** : volume, relation note vs votes
- **Années** : évolution par année / décennie
- **Durées** : distribution, durée moyenne

- **Acteurs** : acteurs les plus représentés (sur un top 3 minimum par film)

Données à collecter

Pour chaque film du Top 250, extraire au minimum :

- **Titre**
- **Année** de sortie
- **Note IMDb** (/10)
- **Nombre de votes**
- **Durée** (minutes)
- **Genre(s)** (multi-valeurs possibles)
- **Réalisateur(s)**
- **Acteurs principaux** (top 3 minimum)
- **Synopsis / description**
- **URL du film**

Fonctionnalités du dashboard (Streamlit)

1) Visualisations obligatoires

Statistiques globales (KPIs)

- Nombre total de films analysés
- Note moyenne du Top 250
- Année moyenne de sortie
- Durée moyenne des films

Graphiques

- Distribution des notes (histogramme)
- Top 10 des films les mieux notés (bar chart)
- Répartition par genre (pie chart ou bar chart)
- Évolution des notes par année (line chart)
- Top 10 des réalisateurs les plus représentés (bar chart)

2) Filtres interactifs

- Filtre **année** (fourchette)
- Filtre **note minimum** (slider)
- Filtre **genre** (selectbox)
- Filtre **réalisateur** (selectbox avec recherche)

3) Tableau de données

- Tableau interactif des films filtrés
- Tri par colonnes (titre, année, note, votes, etc.)

- Limitation d'affichage (ex : 20 films par page)

4) Export

- Bouton pour télécharger les données filtrées en **CSV**

Ressources

- BeautifulSoup Documentation (bs4)
- Streamlit Documentation
- Plotly Documentation