

Thème : le cinéma

URL de la collecte : https://www.imdb.com/search/title/?title_type=feature&sort=alpha,desc

Types de données collectées : genre, dates de sortie, durées, notes

Justification : nous souhaitons faire une analyse de 50 films pour chaque genre triés par ordre alphabétique et ensuite par durée. Notre but sera ensuite de comparer les genres selon certains éléments.

Fiabilité : les données sont issues du site d'IMDB, site sécurisé et web scrappable.

Outils de web scraping choisis : beautifulsoup, requests, matplotlib

Data visualisation prévue :

- graphique à barres (colonnes) -> comparaison de la durée moyenne des films par genre
- graphique à barres (colonnes) -> comparaison de la plus courte/plus longue durée d'un film par genre
- graphique à barres (lignes) -> comparaison de l'étendu des dates de sorties par genre
- graphique à barres (lignes) -> comparaison de la moyenne des notes par genre
- graphique à points -> comparaison des dates les plus récentes/anciennes par genre