# **Ensembles de données IMDb**

Des sous-ensembles de données IMDb sont disponibles pour l'accès aux clients à des fins personnelles et non commerciales. Vous pouvez conserver des copies locales de ces données, et elles sont soumises à nos termes et conditions. Veuillez vous référer aux [licences non commerciales](https://help.imdb.com/article/imdb/general-information/can-i-use-imdb-data-in-my-software/G5JTRESSHJBBHTGX?pf_rd_m=A2FGELUUNOQJNL&pf_rd_p=3aefe545-f8d3-4562-976a-e5eb47d1bb18&pf_rd_r=44N733BW5TH7ZAM9W96Q&pf_rd_s=center-1&pf_rd_t=60601&pf_rd_i=interfaces&ref_=fea_mn_lk1) et [aux droits d'auteur/licence](http://www.imdb.com/Copyright?pf_rd_m=A2FGELUUNOQJNL&pf_rd_p=3aefe545-f8d3-4562-976a-e5eb47d1bb18&pf_rd_r=44N733BW5TH7ZAM9W96Q&pf_rd_s=center-1&pf_rd_t=60601&pf_rd_i=interfaces&ref_=fea_mn_lk2) et vérifier la conformité.

Emplacement

des données Les fichiers de l'ensemble de données peuvent être consultés et téléchargés à partir de <https://datasets.imdbws.com/> . Les données sont actualisées quotidiennement.

Détails de l'ensemble de données IMDb

Chaque ensemble de données est contenu dans un fichier au format TSV (valeurs séparées par des tabulations) compressé dans le jeu de caractères UTF-8. La première ligne de chaque fichier contient des en-têtes décrivant le contenu de chaque colonne. Un *'\N'*est utilisé pour indiquer qu'un champ particulier est manquant ou nul pour ce titre/nom. Les ensembles de données disponibles sont les suivants :

## **title.akas.tsv.gz - Contient les informations suivantes pour les titres :**

* titleId (string) - un tconst, un identifiant alphanumérique unique du titre
* commande (entier) - un nombre pour identifier de manière unique les lignes pour un titleId donné
* title (string) – le titre localisé
* region (string) - la région pour cette version du titre
* langue (chaîne) - la langue du titre
* types (tableau) - Ensemble énuméré d'attributs pour ce titre alternatif. Un ou plusieurs des éléments suivants : "alternative", "dvd", "festival", "tv", "video", "working", "original", "imdbDisplay". De nouvelles valeurs peuvent être ajoutées à l'avenir sans avertissement
* attributs (tableau) - Termes supplémentaires pour décrire ce titre alternatif, non énumérés
* isOriginalTitle (booléen) – 0 : pas le titre d'origine ; 1 : titre d'origine

## **title.basics.tsv.gz - Contient les informations suivantes pour les titres :**

* tconst (string) - identifiant alphanumérique unique du titre
* titleType (chaîne) - le type/format du titre (par exemple, film, court métrage, série télévisée, épisode télévisé, vidéo, etc.)
* primaryTitle (chaîne) - le titre le plus populaire / le titre utilisé par les cinéastes sur le matériel promotionnel au moment de la sortie
* originalTitle (chaîne) - titre original, dans la langue d'origine
* isAdult (booléen) - 0 : titre non adulte ; 1 : titre adulte (filter après)
* startYear (YYYY) – représente l'année de sortie d'un titre. Dans le cas des séries télévisées, il s'agit de l'année de début de la série
* endYear (YYYY) – Année de fin de la série télévisée. '\N' pour tous les autres types de titre
* runtimeMinutes - durée d'exécution principale du titre, en minutes
* genres (string array) – comprend jusqu'à trois genres associés au titre

## **title.crew.tsv.gz - Contient les informations sur le réalisateur et l'écrivain pour tous les titres dans IMDb. Les champs incluent :**

* tconst (string) - identifiant alphanumérique unique du titre
* directeurs (tableau de nconsts) - directeur(s) du titre donné
* écrivains (tableau de nconsts) - écrivain (s) du titre donné

## **title.episode.tsv.gz - Contient les informations sur l'épisode télévisé. Les champs incluent : (exclus)**

* tconst (string) - identifiant alphanumérique de l'épisode
* parentTconst (string) - identifiant alphanumérique de la série télévisée parent
* seasonNumber (entier) - numéro de saison auquel appartient l'épisode
* episodeNumber (entier) - numéro d'épisode du tconst dans la série télévisée

## **title.principals.tsv.gz – Contient la distribution/l'équipe principale pour les titres**

* tconst (string) - identifiant alphanumérique unique du titre
* commande (entier) - un nombre pour identifier de manière unique les lignes pour un titleId donné
* nconst (string) - identifiant alphanumérique unique du nom/de la personne
* catégorie (chaîne) - la catégorie d'emploi dans laquelle la personne se trouvait
* job (chaîne) - le titre du poste spécifique, le cas échéant, sinon '\N'
* caractères (chaîne) - le nom du personnage joué le cas échéant, sinon '\N'

## **title.ratings.tsv.gz - Contient la note IMDb et les informations sur les votes pour les titres**

* tconst (string) - identifiant alphanumérique unique du titre
* averageRating - moyenne pondérée de toutes les évaluations individuelles des utilisateurs
* numVotes - nombre de votes que le titre a reçus

## **name.basics.tsv.gz – Contient les informations suivantes pour les noms :**

* nconst (string) - identifiant alphanumérique unique du nom/de la personne
* primaryName (string)– nom par lequel la personne est le plus souvent créditée
* année de naissance – au format AAAA
* deathYear - au format AAAA si applicable, sinon '\N'
* primaryProfession (tableau de chaînes)– les 3 principales professions de la personne
* knownForTitles (tableau de tconsts) - titres pour lesquels la personne est connue