



1. Présentation des données fournies :

Nous avons reçu de la part de notre client quatre fichiers de type csv notamment **Movies.csv**, **Tags.csv**, **Links.csv**, **ratings.csv**, le tableau ci dessous décrit les fichiers fournis

Fichier	Attributs	Description
Movies	movieId, title, genre	Contient un listing de film, classé par le titre et le genre.
Tags	userId, movieId, tag, timestamp	Contient les tags des utilisateurs identifiés pour des films en fonction de l'heure.
Links	movieId, imdbId, tmdbId	contient des liens numériques servant à compléter des liens url pour trouver des films sur des sites spécialisés.
ratings	userId, movieId, ratings, timestamp	contient les votes données sur des films par certains utilisateurs identifiés, ainsi que la date et l'heure auxquelles l'évaluation a eu lieu.



2. Présentation graphique des données

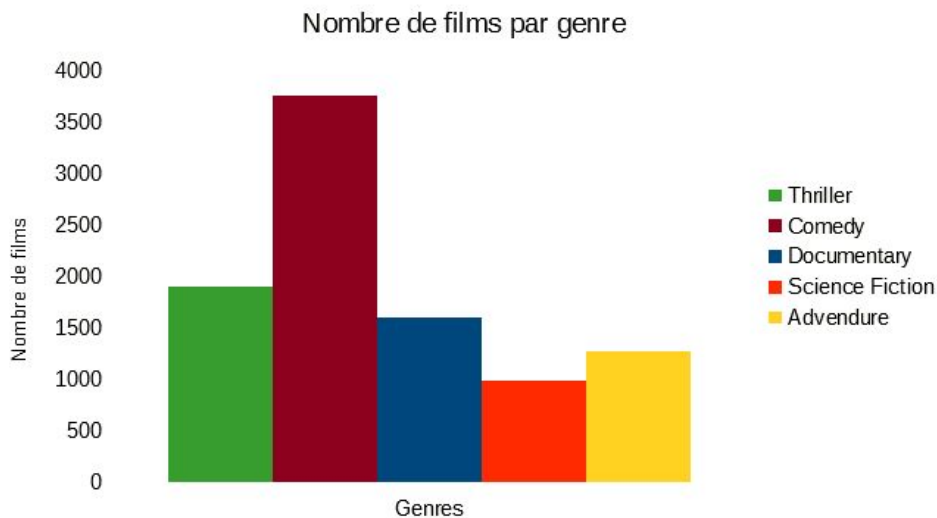


Figure 1: représentation du nombre de film par genre de moviedb

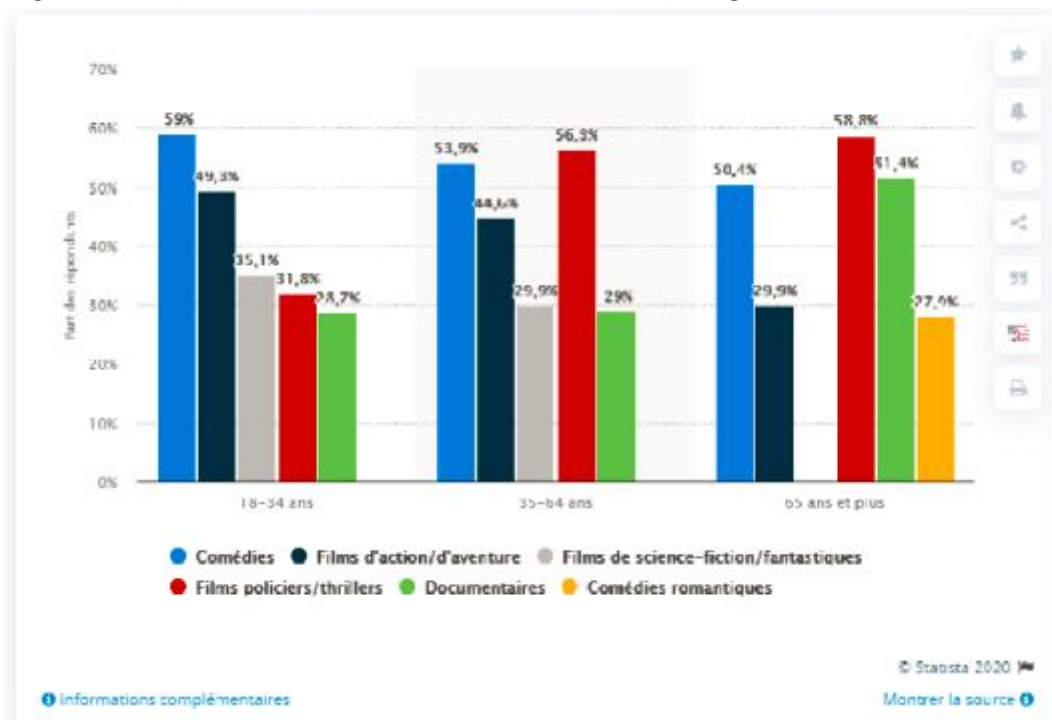


Figure 2: Genres de films préférés des Français en 2016, par tranche d'âge

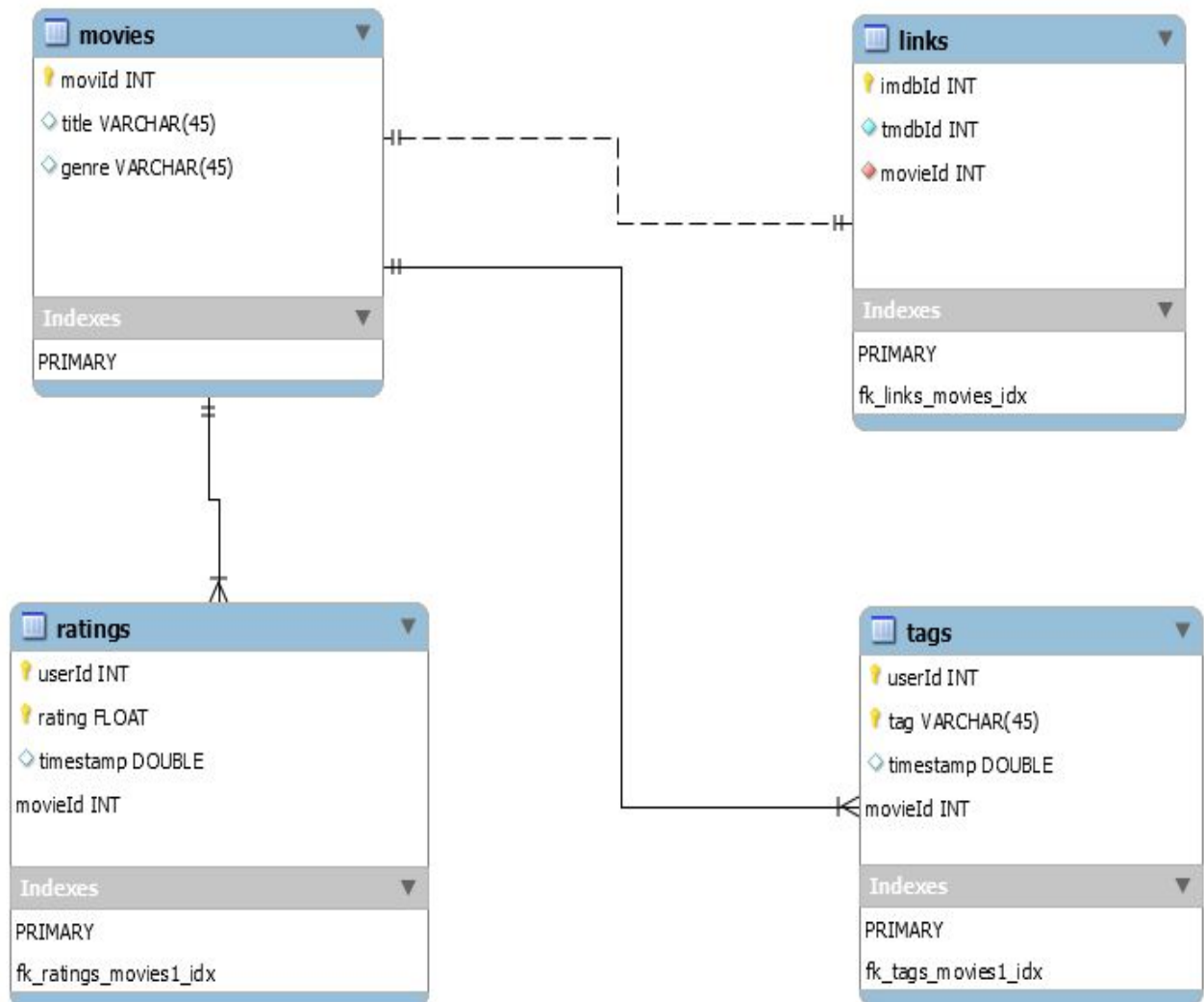
Les genres les plus populaires sont bien représentés dans notre base de données avec un large choix de films. Par exemple, il y a 3756 films catégorisés dans le genre comédie, 1894 pour les Thrillers, 1596 pour les Romances, 1263 films d'aventures, 980 de Science Fiction et 440 documentaires. Et on voit bien en comparaison avec la figure 2, que les genres les plus populaires sont bien présents, tels que les genres comédies et Thriller.



Présentation et analyses des données

3. Schéma relationnel de la base movies :

Le schéma en dessous représente les données telles qu'elles sont fournies par le client ainsi que la relation entre elles.



4. Présentation améliorée des données :

Pour mieux gérer les données nous proposons d'appliquer les modifications notées en dessous:

- Fusionner le tableau `links` avec le tableau `movie.ID`
- Pour le tableau `tags`, on a plusieurs tags pour un film et on le même tag pour différents films. Il faudrait une nouvelle table intermédiaire (`tags - movies`).
- Formater le titre `movies` en séparant le titre de l'année.



5. Analyse poussée (SQL, python) des données de la base:

Pour présenter les données d'une manière un peu poussée; nous avons interrogé la BDD en ajoutant 3 requêtes sql en utilisant python. Pour plus de détails voir le fichier `Finale_Queries.ipynb`

6. Synthèse (Améliorations /propositions):

Nous considérons que vos données sont assez intéressantes pour étudier la faisabilité d'un projet de plateforme en ligne. Par exemple, le tableau ratings permet de dégager les films très bien notés (5.0), la table movie montre beaucoup de genres différents et plus généralement une quantité de film suffisante. Cependant des axes d'améliorations peuvent être mis en place.

Nous pouvons vous proposer, pour éventuellement actualiser l'offre de film, d'identifier les films non présents dans la liste rating (films non notés ou non vus) et de les retirer de la base afin de réactualiser la base. 18 films n'ont pas eu de notes attribuées, on pourrait donc les enlever de notre catalogue car cela peut traduire un désintérêt des utilisateurs pour ces films.

Dans la même idée, en vue d'actualiser l'offre, nous avons fait ressortir la liste de films qui n'apparaissent pas dans les recherches faites par les utilisateurs. En croisant nos bases movies et tags. Il en ressort que 8170 films ne sortent pas dans les recherches. Peut être, il y a t il un problème de termes rattachés au film. Cela voudrait dire que seuls 1572 films de notre base sont recherchés sur un panel de 610 utilisateurs et une base de 9742 films.

Il est important de pouvoir réactualiser le choix des films avec des plus récents. Presque la moitié de nos données, soit environ 3000 films, datent des années 2000-2010. Nous jugeons important l'entrée régulière de nouveau film, pour ne pas lasser l'utilisateur et renouveler notre offre. Dans ce sens, des mise à jour régulière doivent être mise en place

Afin de pouvoir étendre notre offre. Il pourrait être stratégiquement intéressant de développer des genres niche. Par exemple, un des genres les plus appréciés chez les plus de 65 ans sont les documentaires (3ème genre le plus apprécié figure 2). Pour accrocher ces potentiels utilisateurs, on pourrait augmenter l'offre et proposer différents types de documentaires (historique, animalier,...).



Présentation et analyses des données

Enfin, d'un point de vue juridique et dans un souci de respecter la législation concernant les droits d'utilisation de nos données. Il est important que l'on puisse obtenir toutes les licences autorisant l'exploitation des données dans un but commercial et leur mise en ligne.

Merci d'avoir pris en considération notre demande.
Pour toute question, n'hésitez pas à nous contacter.

AKA Team B.