

<b>Introduction</b>	<b>1</b>
The movie database	2
Google, recherches populaires :	3
Les données	4
Base de données the movie database	5
base de données Google	5
Base de données finale	6
<b>I- Requêtes simples sur la base de données</b>	<b>6</b>
A- Interroger la base de données selon les meilleures notes	6
B - Films les plus populaires	7
B- Films avec le plus de vote	8
C - Statistiques sur les données, la popularité moyenne selon les années.	8
D - Le nombre de films par genre	9
E - Filter sur un pays	10
F - Rechercher un film par mot clé dans le titre et le résumé	11
<b>II- requêtes plus complexes</b>	<b>11</b>
A- Nombre de films selon la recherche par pays et par la langue parlée.	11
C- Les genres les plus populaires au cours du temps.	13
D - Filtrage des films à partir de mots clés dans le résumé	14
E - Filtre final pour distinguer les films les plus recherchés et populaires	15
<b>III - Graphs et table de fréquence sur nos données</b>	<b>16</b>
A - Les paires de genre les plus fréquentes	18
B- Les 10 genres les plus connectés	18
C - Filtre des co-occurrences avec action	19
<b>Difficulté :</b>	<b>19</b>
<b>Conclusion</b>	<b>20</b>
<b>Script</b>	<b>21</b>

## Introduction

Le projet que nous allons vous présenter porte sur la thématique des films et des séries. Il s'oriente vers l'analyse des tendances que nous pourrions distinguer à travers les caractéristiques des films, des séries et de l'industrie cinématographique en général. Il s'agit ici du rapport écrit, du contenu de notre script et les

commentaires que nous en tirons. Nous avons posté une vidéo pour montrer l'orchestration de fichier via airflow.

Notre projet consiste à fusionner deux bases de données en lien avec le cinéma, afin d'ajouter une valeur analytique à cette fusion et répondre à des questionnements précis. Nous avons donc sélectionné deux bases de données, que nous allons présenter en expliquant comment nous les avons recherchées et comment nous en avons extrait les informations.

La première base de données concerne les films. Chaque utilisateur peut y ajouter un film en renseignant son résumé, sa date de sortie et ses auteurs. Il s'agit d'une plateforme en open source, qui permet d'extraire un maximum d'informations via une API.

Nous avons ensuite intégré une seconde base de données, plus technique, qui apporte des informations complémentaires aux caractéristiques classiques des films (genre, date de sortie, réalisateur, acteurs, etc.). Cette base de données est issue des requêtes les plus utilisées dans la barre de recherche Google. Nous avons choisi cette source car il nous était techniquement difficile d'effectuer du scraping ou d'interroger directement les API de Google. Cette base de données nous a semblé être un bon compromis pour enrichir notre analyse.

Dans la suite du projet, nous détaillerons ces deux bases de données, puis nous expliquerons les manipulations réalisées pour optimiser notre base de données finale. Enfin, nous présenterons les différentes requêtes effectuées afin d'analyser le marché du cinéma et d'identifier les tendances qui s'en dégagent.

Nous avons utilisé plusieurs logiciels pour coder, notamment Google Colab et Visual Studio Code. Certaines bibliothèques, comme PySpark, n'étaient pas compatibles avec nos machines sur Visual Studio Code, ce qui nous a amenés à privilégier Google Colab pour certaines étapes. Toutefois, le rendu final a été réalisé sur Visual Studio Code.

Les captures d'écran présentées dans notre rapport proviennent principalement de Google Colab, car elles sont plus lisibles que celles issues des dataframes sous Pandas. L'objectif final de notre projet, la création d'un DAG (\*Directed Acyclic Graph\*), a été rendu possible grâce à Visual Studio Code et à la bibliothèque Pandas.

Enfin, notre projet est entièrement reproductible par un tiers, à l'exception de la base de données issue de Google, que nous avons dû filtrer manuellement.

## The movie database

The Movie Database est une plateforme collaborative qui fournit une base de données sur les films et séries, incluant des informations détaillées comme les résumés, les notes, les dates de sortie et les castings.

La première base de données que nous avons choisie provient d'une extraction de données via une requête effectuée sur le site \*The Movie Database\*. Nous avons décidé d'extraire les 500 premières pages de la section des films les plus populaires. Cette extraction permet d'obtenir des informations mises à jour en permanence, à condition d'actualiser notre requête Python.

Nous avons également testé d'autres requêtes afin d'extraire davantage d'informations. Par exemple, nous avons tenté d'extraire les films sortis entre 1900 et 2024, mais cette opération était très coûteuse en ressources, nécessitant plus de 50 minutes d'exécution sur les machines de Nanterre.

Suite à l'extraction via l'API, nous avons obtenu une base de données de 10 084 lignes, un nombre qui peut varier en fonction de l'actualisation des données sur le site. Cette base est composée de neuf colonnes : le titre, la date de sortie, la note, le résumé, le nombre de votes, la popularité, le titre original, la langue originale et les genres.

Une ligne type est comme par exemple :

Titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	Original Title	Original Language	Genres
Venom: The Last D...	2024-10-22	6.6	Eddie et Venom so...	1113	12656.263	Venom: The Last D...	en	Action, Science-F...

## Google, recherches populaires :

Nous avons extrait une base de données sur kaggle qui regroupe les mots-clés les plus recherchés dans la base de recherche de Google entre 2008 et 2020.

Kaggle est une plateforme en ligne qui propose des bases de données, des compétitions en science des données et un environnement de travail collaboratif pour l'analyse et le machine learning.

Cette base de données regroupe un grand nombre d'informations qui ne sont pas exclusivement liées aux films. Elle contient par exemple des mots-clés associés à l'actualité, aux pays, aux personnalités, aux sports ou encore aux événements. Nous avons donc effectué un filtrage manuel des données, car les colonnes du fichier brut étaient difficiles à manipuler. L'extraction était compliquée en raison de la présence de plusieurs colonnes identiques, mais dans différentes langues, ainsi que de catégories légèrement différentes, comme "film", "movies" ou "popular movies". Nous avons donc réalisé un tri manuel pour sélectionner les colonnes les plus pertinentes afin de conserver un maximum d'informations sur les catégories de films, qu'il s'agisse de séries, de films ou du cinéma en général. Les termes que nous avons retenus correspondent aux sujets les plus recherchés dans la barre de recherche de Google entre 2008 et 2020. Par exemple, "Venom II" est associé à ces

recherches. Si ce mot-clé a été fortement recherché au cours d'une année, il peut apparaître dans le top 5 des requêtes les plus effectuées pour une année et un sujet donné sur Google.

Cette base de données contient cinq colonnes et 880 lignes. Les colonnes incluent le pays (avec une entrée spécifique pour l'ensemble des pays regroupés), la date de la recherche, la catégorie correspondant au sujet de la recherche, le classement allant de un à cinq selon la fréquence des recherches et enfin "query", qui représente le nom du film recherché.

Country	Year	Category	Rank	Query
Global	2001	Movies	1	Harry Potter

## Les données

La base de données contient plusieurs types de données, notamment des données numériques et catégorielles. Afin d'optimiser notre analyse, nous avons procédé à la jointure des deux bases de données.

Avant d'effectuer cette fusion, nous avons modifié certains éléments dans la base de données des films. Nous avons ajusté les titres et les titres originaux pour qu'ils correspondent mieux aux noms présents dans la seconde base de données, celle issue de Google, afin que la jointure capture le maximum d'informations. Cette fusion a été réalisée en reliant la colonne "query" de la base Google à la colonne "Titre" de la base de films.

Grâce à cette jointure, nous avons obtenu une nouvelle base de données qui permet d'observer, pour chaque film, son apparition dans The Movie Database et son niveau de recherche sur Google selon les années et les pays. Cela nous permet de formuler des hypothèses, comme vérifier si la popularité d'un film est bien corrélée au nombre de recherches effectuées sur Internet. Il est également possible d'analyser les différences de tendances cinématographiques selon les pays.

Nous avons filtré notre dataframe en ne conservant que les films sortis entre 2008 et 2020, car la base de données Google ne couvre que cette période. Par conséquent, notre dataframe est passé d'environ 10 000 lignes à près de 4 000 lignes.

Cependant, ce nombre peut évoluer dans le temps, car il dépend des mises à jour de la base Google et des sorties de films. Ainsi, notre base de données est dynamique et s'adapte en fonction de l'actualisation des fichiers sources.

Nous avons choisi de conserver l'extraction initiale issue de The Movie Database, car nous pensons qu'il serait techniquement possible d'effectuer un scraping sur Google pour obtenir davantage d'informations. Toutefois, nous ne disposons pas des ressources nécessaires pour réaliser cette extraction de manière automatique. La taille finale de notre base de données peut légèrement varier selon la date de l'extraction, mais elle contient en général environ 10 000 lignes. Seule la base Google reste stable avec 880 lignes. Nous présentons ci-dessous un aperçu des deux dataframes initiaux ainsi que du dataframe final obtenu après leur fusion.

Base de données the movie database

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres
Venom: The Last D...	2024-10-22	6.6	Eddie et Venom so...	1113	12656.263	Venom: The Last D...	en	Action, Science-F...
Vaiana 2	2024-11-21	6.864	Vaiana 2 embarque...	420	4485.032	Moana 2	en	Animation, Aventu...
Elevation	2024-11-07	6.4	Les survivants d'...	126	2554.883	Elevation	en	Action, Science-F...
Absolution	2024-10-31	6.1	Ancien boxeur, Th...	120	1846.723	Absolution	en	Action, Crime, Th...
Watchmen: Chapter II	2024-11-25	7.738	Soupçonnant les é...	42	1439.232	Watchmen: Chapter II	en	Animation, Mystèr...
Armor	2024-10-30	5.6	Un père et son fi...	54	1423.925	Armor	en	Action, Crime, Th...
Gladiator II	2024-11-05	6.7	Des années après ...	1006	1262.621	Gladiator II	en	Action, Aventure, ...
Sidelined: The QB...	2024-11-29	6.6	NULL	58	1338.291	Sidelined: The QB...	en	Comédie, Romance, ...
Miraculous World ...	2024-11-14	8.71	Afin de sauver l'...	31	1219.844	Miraculous World ...	fr	Animation, Aventu...
Terrifier 3	2024-10-09	6.85	Après avoir survé...	1201	922.731	Terrifier 3	en	Horreur, Thriller
Apocalipsis Z: el...	2024-10-04	6.86	Lorsqu'une sorte ...	677	958.51	Apocalypse Z : le...	es	Drame, Action, Ho...
Wicked	2024-11-20	7.597	Elphaba, une jeun...	485	1084.787	Wicked	en	Drame, Romance, F...
Alien : Romulus	2024-08-13	7.2	Quelques années a...	2588	917.605	Alien: Romulus	en	Horreur, Science-...
Deadpool & Wolverine	2024-07-24	7.7	Un Wade Wilson dé...	5839	893.759	Deadpool & Wolverine	en	Action, Comédie, ...
Smile 2	2024-10-16	6.6	Sur le point d'en...	888	970.057	Smile 2	en	Horreur, Mystère
The Substance	2024-09-07	7.256	Elisabeth Sparkle...	2475	900.139	The Substance	en	Horreur, Science-...
Red One	2024-10-31	6.6	Après l'enlèvemen...	227	864.206	Red One	en	Action, Comédie, ...
Vice-versa 2	2024-06-11	7.6	Riley, adolescent...	4935	740.187	Inside Out 2	en	Animation, Aventu...
Levels	2024-11-01	5.689	Après avoir été t...	37	792.543	Levels	en	Action, Science-F...
Transformers : Le...	2024-09-11	8.1	Les origines d'Op...	843	725.848	Transformers One	en	Animation, Scienc...

only showing top 20 rows

10084

base de données Google

Country	Year	Category	Rank	Query
Global	2001	Movies	1	Harry Potter
Global	2001	Movies	2	Lord of the Rings
Global	2001	Movies	3	Final Fantasy
Global	2001	Movies	4	Tomb Raider
Global	2001	Movies	5	Shrek
Global	2002	Movies	1	Spiderman
Global	2002	Movies	2	Harry Potter
Global	2002	Movies	3	Star Wars
Global	2002	Movies	4	Jackass
Global	2002	Movies	5	Scooby Doo
Australia	2003	Popular Movies	1	2 Fast 2 Furious
Australia	2003	Popular Movies	2	Lord of the Rings
Australia	2003	Popular Movies	3	Finding Nemo
Australia	2003	Popular Movies	4	Texas Chainsaw Ma...
Australia	2003	Popular Movies	5	Pirates of the Ca...
Germany	2003	Beliebte Filme	1	Matrix
Germany	2003	Beliebte Filme	2	Harry Potter
Germany	2003	Beliebte Filme	3	Herr Der Ringe
Germany	2003	Beliebte Filme	4	Findet Nemo
Germany	2003	Beliebte Filme	5	Kill Bill

only showing top 20 rows

880

## Base de données finale

DataFrame filtré (années 2009-2020) :

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
Le Grinch	2018-11-08	6.864	Terré dans sa cav...	3944	477.459	The Grinch	en	Animation, Famili...	NULL	NULL	NULL	NULL	NULL
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	United Kingdom	2018	Movies	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Singapore	2018	Top Trending Movies	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Romania	2018	Filme	3	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Nigeria	2018	Movies	4	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Malaysia	2018	Filme	3	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Hungary	2018	Filme	1	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	France	2018	Les films	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Canada	2018	Movies	4	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Brazil	2018	Filmes	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	Global	2018	Movies	3	Venom
일본 영화	2017-02-09	6.2	NULL	28	364.704	일본 영화	ko	Romance, Drame	NULL	NULL	NULL	NULL	NULL
Des jours plus be...	2010-02-19	6.523	Win et Floor, un ...	22	299.557	Des jours plus be...	nl	Romance	NULL	NULL	NULL	NULL	NULL
Avengers : Infini...	2018-04-25	8.2	Les Avengers et l...	29791	342.224	Avengers: Infini...	en	Aventure, Action,...	NULL	NULL	NULL	NULL	NULL
Equalizer	2014-09-24	7.281	McCall, un homme ...	9029	192.911	The Equalizer	en	Thriller, Action,...	NULL	NULL	NULL	NULL	NULL
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760	238.105	Interstellar	en	Aventure, Drame, ...	United States	2014	Movies	2	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760	238.105	Interstellar	en	Aventure, Drame, ...	United Arab Emirates	2014	Movies	2	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760	238.105	Interstellar	en	Aventure, Drame, ...	Sweden	2014	Film	1	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760	238.105	Interstellar	en	Aventure, Drame, ...	Singapore	2014	Movies	1	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760	238.105	Interstellar	en	Aventure, Drame, ...	Malaysia	2014	Movies (Film)	3	Interstellar

only showing top 20 rows

4877

Nous allons donc pouvoir maintenant entamer les requêtes afin d'interroger notre base de données.

## I- Requêtes simples sur la base de données

Nous allons effectuer des requêtes de plus en plus précises afin d'explorer les différentes possibilités d'interrogation de cette base de données. Au fur et à mesure, nous constaterons que se limiter à des requêtes basées uniquement sur des conditions simples n'est pas toujours la méthode la plus pertinente.

L'objectif est de présenter différentes requêtes et d'analyser lesquelles permettent de mieux comprendre notre base de données tout en extrayant les hypothèses les plus pertinentes. Parmi les questions que nous chercherons à éclaircir, nous tenterons de déterminer quels sont les films les plus populaires de ces dix dernières années, quel est le genre prédominant au cours de cette période et quel pays apparaît comme le plus prolifique en tant que producteur de films selon les données dont nous disposons.

### A- Interroger la base de données selon les meilleures notes

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
여자 하숙집 3]	2019-06-11]	9.8]	NULL]	2.0]	21.277]	여자 하숙집 3]	ko]	Romance]	NULL]	NULL]	NULL]	NULL]	NULL]
働く女の艶めかしい完全着衣ファック...	2016-12-13]	9.5]	NULL]	2.0]	38.398]	働く女の艶めかしい完全着衣ファック...	ja]	Drame]	NULL]	NULL]	NULL]	NULL]	NULL]
Sree Hari Hara Su...	2017-04-17]	9.0]	NULL]	3.0]	38.932]	Sree Hari Hara Su...	ja]	ta]	Animation, Drame,...	NULL]	NULL]	NULL]	NULL]
ネトラセラレ3 ～愛～	2018-04-02]	9.0]	NULL]	1.0]	37.372]	ネトラセラレ3 ～愛～	ja]	Drame]	NULL]	NULL]	NULL]	NULL]	NULL]
Moananuiākea: One...	2018-11-18]	9.0]	NULL]	1.0]	17.864]	Moananuiākea: One...	en]	Documentaire]	NULL]	NULL]	NULL]	NULL]	NULL]
ハイキュー!! 終わりと始まり]	2015-07-03]	8.6]	Premier Haikyuu !...]	36.0]	13.696]	Haikyuu !! - Film ...]	ja]	Animation, Drame]	NULL]	NULL]	NULL]	NULL]	NULL]
The Dark Knight : ...]	2008-07-16]	8.516]	Batman aborde une...]	32971.0]	142.229]	The Dark Knight]	en]	Drame, Action, Cr...	NULL]	NULL]	NULL]	NULL]	NULL]
기생충]	2019-05-30]	8.506]	Toute la famille ...]	18347.0]	110.555]	Parasite]	ko]	Comédie, Thriller...	NULL]	NULL]	NULL]	NULL]	NULL]
동행회의 목적: 더 비가림]	2017-02-02]	8.5]	NULL]	6.0]	75.156]	동행회의 목적: 더 비가림]	ko]	Comédie, Romance]	NULL]	NULL]	NULL]	NULL]	NULL]
君の名は。]	2016-08-26]	8.485]	Mitsuha est une l...	11406.0]	118.19]	Your Name.]	ja]	Animation, Romanc...	NULL]	NULL]	NULL]	NULL]	NULL]

On observe que les films les mieux notés ont des scores proches de 10, avec un premier film noté 9,8, qui est un film coréen. Cependant, le nombre de votes reste très faible pour l'ensemble des films, ce qui limite la pertinence de cette méthode pour identifier les films les plus populaires en fonction des votes. Ce filtrage n'est donc pas le plus optimisé pour déterminer la popularité réelle des films, car un film peut avoir une très bonne note tout en étant peu regardé.

## B - Films les plus populaires

DataFrame filtré (années 2008-2020) :

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
Le Grinch]	2018-11-08]	6.864]	Terré dans sa cav...	3944]	477.459]	The Grinch]	en]	Animation, Famili...	NULL]	NULL]	NULL]	NULL]	NULL]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	United Kingdom]	2018]	Movies]	5]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Singapore]	2018]	Top Trending Movies]	5]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Romania]	2018]	Filme]	3]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Nigeria]	2018]	Movies]	4]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Malaysia]	2018]	Filme]	3]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Hungary]	2018]	Filme]	1]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	France]	2018]	Les films]	5]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Canada]	2018]	Movies]	4]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Brazil]	2018]	Filmes]	5]	Venom]
Venom]	2018-09-28]	6.8]	Des symbiotes déb...	16065]	301.84]	Venom]	en]	Science-Fiction, ...	Global]	2018]	Movies]	3]	Venom]
일본 영화]	2017-02-09]	6.2]	NULL]	28]	364.784]	일본 영화]	ko]	Romance, Drame]	NULL]	NULL]	NULL]	NULL]	NULL]
Des jours plus be...	2010-02-19]	6.523]	Win et Floor, un ...]	22]	299.557]	Des jours plus be...	n]	Romance]	NULL]	NULL]	NULL]	NULL]	NULL]
Avengers : Infini...	2018-04-25]	8.2]	Les Avengers et l...	29791]	342.224]	Avengers: Infini...	en]	Aventure, Action,...	NULL]	NULL]	NULL]	NULL]	NULL]
Equalizer]	2014-09-24]	7.281]	McCall, un homme ...]	9029]	192.911]	The Equalizer]	en]	Thriller, Action,...	NULL]	NULL]	NULL]	NULL]	NULL]
Interstellar]	2014-11-05]	8.4]	Dans un futur pro...	35760]	238.105]	Interstellar]	en]	Aventure, Drame, ...	United States]	2014]	Movies]	2]	Interstellar]
Interstellar]	2014-11-05]	8.4]	Dans un futur pro...	35760]	238.105]	Interstellar]	en]	Aventure, Drame, ...	United Arab Emirates]	2014]	Movies]	2]	Interstellar]
Interstellar]	2014-11-05]	8.4]	Dans un futur pro...	35760]	238.105]	Interstellar]	en]	Aventure, Drame, ...	Sweden]	2014]	Film]	1]	Interstellar]
Interstellar]	2014-11-05]	8.4]	Dans un futur pro...	35760]	238.105]	Interstellar]	en]	Aventure, Drame, ...	Singapore]	2014]	Movies]	1]	Interstellar]
Interstellar]	2014-11-05]	8.4]	Dans un futur pro...	35760]	238.105]	Interstellar]	en]	Aventure, Drame, ...	Malaysia]	2014]	Movies (Film)]	3]	Interstellar]

only showing top 20 rows

4077

Cette requête permet d'observer les films les plus populaires en se basant sur la colonne de popularité. On constate que le film Grinch obtient un score très élevé de 476. Cependant, il ne figure pas dans la base de données Google, car son titre n'apparaît pas dans la colonne query.

On remarque également des doublons dans la colonne Titre pour le film Venom. Ces doublons sont intéressants, car ils permettent de voir, grâce à la colonne query, que ce film a été populaire dans plusieurs pays au cours de la même année, notamment à Singapour, au Nigeria, au Canada et au Brésil.

De plus, au niveau global, Venom se classe comme le troisième film le plus recherché sur la barre de recherche de Google. À partir de cette requête simple, on peut donc affirmer que pour l'année 2018, Venom était un film extrêmement populaire.

## B- Films avec le plus de vote

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
Inception	2010-07-15	8.369	Dom Cobb est un v...	36635.0	122.2	Inception	en	Action, Science-F...	NULL	NULL	NULL	NULL	NULL
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	Canada	2014	Movies	2	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	Malaysia	2014	Movies (Film)	3	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	Ireland	2014	Movies	5	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	United Arab Emirates	2014	Movies	2	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	Sweden	2014	Film	1	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	Singapore	2014	Movies	1	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	United States	2014	Movies	2	Interstellar
Interstellar	2014-11-05	8.4	Dans un futur pro...	35760.0	238.105	Interstellar	en	Aventure, Drame, ...	France	2014	Films	2	Interstellar
The Dark Knight : ...	2008-07-16	8.516	Batman aborde une...	32971.0	142.229	The Dark Knight	en	Drame, Action, Cr...	NULL	NULL	NULL	NULL	NULL

En utilisant uniquement le nombre de votes comme filtre, on constate que le film Interstellar a reçu un très grand nombre de votes, tout comme Inception. En croisant ces données avec la base de données Google, on observe que Inception est un film extrêmement populaire, apprécié sur plusieurs continents, notamment en Europe avec la Suède, en Asie avec Singapour et en Amérique avec les États-Unis.

De plus, on remarque que ce film présente à la fois une popularité élevée et un grand nombre de votes, ce qui renforce son statut de film largement visionné et recherché. Ces observations suggèrent qu'il serait pertinent d'affiner l'analyse en combinant plusieurs filtres afin d'obtenir une vision plus complète de la popularité des films.

## C - Statistiques sur les données, la popularité moyenne selon les années.

Year	avg(Popularité)
2008	47.482321126837476
2009	38.556854092512005
2010	123.13161097809859
2011	43.98138288979058
2012	37.12952691252514
2013	35.5063122799522
2014	52.13135466752229
2015	60.270837715861255
2016	43.484417385243354
2017	38.310180148807184
2018	49.97792479261082
2019	47.21958457897236
2020	37.49174136457772



Ici, nous observons la moyenne de la colonne popularité en fonction des années. Cette analyse permet de constater que certains films ont été plus populaires que d'autres selon les périodes. On remarque que l'année où le plus grand nombre de films populaires a été produit est 2010, avec une moyenne élevée de 123, tandis que 2013 apparaît comme une année moins marquante, avec une moyenne de 35.

Cette statistique pourrait être mise en relation avec une étude plus large sur l'industrie du cinéma afin de comprendre quelles années ont été les plus propices à la production de films à grand succès. Une analyse approfondie pourrait ainsi révéler des tendances dans la production cinématographique et leur lien avec les habitudes du public.

## D - Le nombre de films par genre

Genre	count
Drame	787
Thriller	771
Drame	752
Action	688
Aventure	640
Comédie	605
Comédie	495
Action	461
Romance	444
Fantastique	444
Science-Fiction	417
Familial	378
Horreur	312
Crime	310
Animation	302
Mystère	267
Horreur	239
Thriller	229
Romance	206
Animation	197

only showing top 20 rows

Ce tableau permet d'afficher le nombre de films par genre. On constate la présence de doublons, comme "Drame" et "Comédie" ou encore "Horreur". Cette situation est due à la structure de notre base de données, où les genres apparaissent plusieurs fois dans différentes colonnes en fonction de leur position dans les observations. Par exemple, le mot "Comédie" peut apparaître 650 fois en première position et 495 fois en deuxième position. En les cumulant, on obtient un total plus représentatif du nombre de films correspondant à ce genre.

On observe ainsi que les films de "drame, thriller et comédie" figurent parmi les plus représentés, chacun dépassant les 1 000 occurrences dans la base de données.

Ces genres semblent donc être les plus fréquemment produits et recherchés dans l'industrie cinématographique.

## E - Filter sur un pays

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
Venom	2018-09-28	6.8	Des symbiotes déb...	16065	301.84	Venom	en	Science-Fiction, ...	United Kingdom	2018	Movies	5	Venom
Joker	2019-10-01	8.1	Dans les années 1...	25781	201.229	Joker	en	Crime, Thriller, ...	United Kingdom	2019	Movies	2	Joker
Deadpool	2016-02-09	7.624	Deadpool, est l'a...	31055	190.413	Deadpool	en	Action, Aventure, ...	United Kingdom	2016	Movies	1	Deadpool
American Sniper	2014-12-25	7.426	Tireur d'élite de...	12815	168.805	American Sniper	en	Guerre, Action	United Kingdom	2015	Films	5	American Sniper
Black Panther	2018-02-13	7.382	Après les événeme...	22164	138.884	Black Panther	en	Action, Aventure, ...	United Kingdom	2018	Movies	1	Black Panther
Deadpool 2	2018-05-10	7.492	Deadpool se voit ...	17954	143.977	Deadpool 2	en	Action, Comédie, ...	United Kingdom	2018	Movies	4	Deadpool 2
World War Z	2013-06-19	6.818	Gerry Lane se lan...	15342	81.861	World War Z	en	Action, Horreur, ...	United Kingdom	2013	Movies	2	World War Z
The Revenant	2015-12-25	7.533	Dans une Amérique...	18053	106.682	The Revenant	en	Western, Drame, A...	United Kingdom	2016	Movies	3	The Revenant
Jurassic World	2015-06-06	6.692	Après le décès de...	20315	69.709	Jurassic World	en	Action, Aventure, ...	United Kingdom	2015	Films	1	Jurassic World
Iron Man 3	2013-04-18	6.931	Tony Stark, l'ind...	22142	89.501	Iron Man 3	en	Action, Aventure, ...	United Kingdom	2013	Movies	3	Iron Man 3
Suicide Squad	2016-08-03	5.9	C'est tellement j...	21093	75.705	Suicide Squad	en	Action, Aventure, ...	United Kingdom	2016	Movies	2	Suicide Squad
Spectre	2015-10-26	6.5	Un message crypti...	10480	68.839	Spectre	en	Action, Aventure, ...	United Kingdom	2015	Films	4	Spectre
Man of Steel	2013-06-12	6.6	Un petit garçon d...	15094	45.167	Man of Steel	en	Action, Aventure, ...	United Kingdom	2013	Movies	1	Man of Steel
Captain Marvel	2019-03-06	6.8	Captain Marvel ra...	15685	58.773	Captain Marvel	en	Action, Aventure, ...	United Kingdom	2019	Movies	3	Captain Marvel
Joker	2015-01-14	5.668	Nick Wild, ex-mar...	1768	40.869	Wild Card	en	Thriller, Crime, ...	United Kingdom	2019	Movies	2	Joker
1917	2019-12-25	7.991	Pris dans la tour...	12378	35.96	1917	en	Guerre, Histoire, ...	United Kingdom	2020	Movies	1	1917
The Gentlemen	2020-01-01	7.68	Quand Mickey Pear...	5824	37.536	The Gentlemen	en	Action, Crime, Co...	United Kingdom	2020	Movies	3	The Gentlemen
Jojo Rabbit	2019-10-18	8.017	Jojo est un petit...	9527	31.174	Jojo Rabbit	en	Comédie, Guerre, ...	United Kingdom	2020	Movies	5	Jojo Rabbit
True Grit	2010-12-22	7.306	1870, juste après...	5232	24.246	True Grit	en	Drame, Aventure, ...	United Kingdom	2011	Movies	4	True Grit
Lion	2016-11-24	8.068	Une incroyable hi...	6533	25.71	Lion	en	Drame	United Kingdom	2020	Movies	4	Lion

Filterer les données sur un seul pays permet d'observer quels sont les films les plus populaires dans cette région. On constate qu'une grande majorité des films les plus recherchés appartiennent aux genres action et aventure, qui dominent le top cinq des recherches Google selon les différentes années.

On remarque également que les films de super-héros sont particulièrement appréciés, avec des titres comme Captain Marvel, Man of Steel, Iron Man, Deadpool et Venom, qui figurent parmi les films les plus recherchés. Cette tendance confirme l'engouement du public pour ce type de productions dans ce pays.

## F - Rechercher un film par mot clé dans le titre et le résumé

```
keyword_movies = spark_dikk.filtered.filter(
    (col("titre").contains("Venom")) | (col("Résumé").contains("symbiotes"))
)
keyword_movies.show()
```

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	United Kingdom	2018	Movies	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Singapore	2018	Top Trending Movies	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Romania	2018	Filme	3	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Nigeria	2018	Movies	4	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Malaysia	2018	Filme	3	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Hungary	2018	Filmek	1	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	France	2018	Les films	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Canada	2018	Movies	4	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Brazil	2018	Filmes	5	Venom
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84	Venom	en	Science-Fiction, ...	Global	2018	Movies	3	Venom

Cette requête apporte une valeur ajoutée en permettant d'effectuer des recherches précises sur un mot-clé spécifique. Il est ainsi possible d'identifier les films dont les résumés et les titres contiennent ce mot.

Il serait également envisageable d'utiliser des termes plus larges, comme "amour", afin d'observer quels films y sont associés. Cette approche permet d'affiner l'analyse en identifiant les thèmes récurrents dans les résumés et en comprenant comment certains mots-clés sont liés à des genres cinématographiques spécifiques.

## II- requêtes plus complexes

Nous allons maintenant effectuer des requêtes plus complexes afin d'approfondir l'analyse de notre dataframe et d'identifier les tendances qui en ressortent. L'objectif est d'obtenir des résultats aussi précis et pertinents que possible, en affinant les critères d'interrogation pour mieux comprendre les dynamiques qui influencent la popularité des films.

A- Nombre de films selon la recherche par pays et par la langue parlée.

Country	count
NULL	3756
Global	29
France	27
Singapore	21
Ireland	21
United Kingdom	20
United States	18
Malaysia	15
Canada	14
Sweden	13

only showing top 10 rows

Cette requête permet d'afficher les films recherchés selon les pays. Bien qu'elle puisse sembler anodine ou peu pertinente à première vue, elle met en évidence le nombre d'observations par pays, notamment pour la catégorie "globale" et pour la France, qui compte environ 30 occurrences.

Cette analyse permet surtout de confirmer que les recherches ont été correctement filtrées dans la base de données Google. En effet, nous avons sélectionné dans la colonne "query" uniquement des termes en lien avec le cinéma, comme "film", "série" ou "movies". Cependant, certaines informations ont pu être omises, car nous ne maîtrisons pas toutes les langues et certaines données étaient écrites dans des alphabets différents. Malgré cela, nous avons tenté de capturer le maximum d'informations pertinentes.

La requête suivante permettra d'avoir une meilleure compréhension de l'origine des films. Bien que nous ne puissions pas, avec cette base de données, retracer précisément le pays de production de chaque film, nous pouvons néanmoins l'associer à une culture spécifique en nous basant sur la langue parlée dans le film.

Original Language	count
en	3008
ko	207
ja	201
fr	134
es	96
zh	45
de	33
NULL	29
hi	25
ru	23
it	17
nl	16
cn	14
da	14
pt	10
th	9
no	9
tl	8
id	8
sv	7

only showing top 20 rows

Nous pouvons observer ici qu'une grande majorité des films sont produits en anglais, tandis que les autres langues restent minoritaires. Dans le top trois, on retrouve l'anglais, suivi du coréen et du japonais.

Nous avons tenté d'inclure le plus grand nombre possible de films issus de différentes régions du monde, bien que certaines langues restent sous-représentées. Par exemple, l'hindi compte seulement 25 films et le russe 23. Cette distribution reflète la domination de certaines industries cinématographiques, tout en mettant en évidence la présence plus limitée de films issus d'autres cultures dans notre base de données.

## C- Les genres les plus populaires au cours du temps.

Year	Genre	Avg_Popularity
2008	en	1797.5
2008	ja	96.0
2008	Science-Fiction	36.11454545454544
2008	Animation	35.304
2008	Science-Fiction	35.2205
2008	Action	33.18409999999999
2008	Action	33.053181818181805
2008	Fantastique	32.91975
2008	Aventure	32.34472413793104
2008	Thriller	31.952621621621624
2008	Guerre	31.072999999999997
2008	Crime	30.517444444444443
2008	Animation	30.167888888888886
2008	Romance	29.69959375
2008	Familial	29.29968965517241
2008	Mystère	29.023374999999998
2008	Drame	28.817928571428574
2008	Aventure	28.365466666666666
2008	Histoire	26.889857142857142
2008	Fantastique	26.451041666666667

only showing top 20 rows

Ici, nous pouvons observer l'évolution de la popularité des différents genres de films au fil des années. On remarque que, dans la colonne des genres, certaines entrées correspondent à des langues parlées plutôt qu'à des catégories cinématographiques. En effet, certains films ont pour particularité d'inclure la langue comme une catégorie.

Nous avons choisi de ne pas effectuer un nettoyage trop strict de ces données afin d'éviter une perte d'information et de préserver au mieux les données issues des deux bases fusionnées. Cette approche permet de conserver un maximum de détails pour mieux comprendre les tendances globales du marché cinématographique.

Les langues pourraient d'ailleurs jouer un rôle clé dans la classification des films et leur réception selon les régions. Malgré un travail de nettoyage portant principalement sur le formatage des données, des doublons subsistent. Par exemple, le genre "science-fiction" apparaît parfois sous différentes formes en raison d'un espace en trop ou d'un caractère inconnu mal interprété par la base de données.

## D - Filtrage des films à partir de mots clés dans le résumé

```
# Filtrer les films avec des résumés contenant des mots-clés
keyword_filtered = spark_dikk_filtered.filter(
    ... col("Résumé").rlike("(?i)aventure|magie|action")
)
keyword_filtered.show(truncate=False)
```

titre	Date de sortie	Note	Résumé
Vaiana, la légende du bout du monde	2016-10-13	7.569	Il y a 3 000 ans, les plus grands marins du monde parcouraient
Les Chroniques de Noël	2018-11-22	7.145	Kate et Teddy sont tout occupés à épier la venue du père Noël.
Dragons 2	2014-06-05	7.7	Tandis qu'Aspid, Rustik et le reste de la bande se défient du
Harry Potter et les Reliques de la mort - 1ère partie	2010-11-17	7.747	Le pouvoir de Voldemort s'étend. Celui-ci contrôle maintenant
La Reine des neiges	2013-11-20	7.245	Anna, une jeune fille aussi audacieuse qu'optimiste, se lance
Pirates des Caraïbes : La Fontaine de jouvence	2011-05-15	6.554	Dans cette histoire pleine d'action, où vérité, trahison, jeune
Inception	2010-07-15	8.369	Dom Cobb est un voleur expérimenté, le meilleur dans l'art d'an
WALL·E	2008-06-22	8.1	Faites la connaissance de WALL·E (prononcez « Walli ») : WALL·E
Rio	2011-04-03	6.7	Blu, un perroquet bleu d'une espèce très rare, quitte sa petite
Dragons : Retrouvailles	2019-12-03	8.1	Quand il devient évident que la nouvelle génération de Vikings
Raïponce	2010-11-24	7.6	Quand le bandit le plus recherché du royaume, Flynn Rider, se
La Princesse et la Grenouille	2009-12-08	7.2	Cette version moderne d'un conte classique par Walt Disney Anir
Spider-Man : New Generation	2018-12-06	8.399	Spider-Man : New Generation suit les aventures de Miles Morales
La-haut	2009-05-28	7.958	À 78 ans, Carl n'envisage plus qu'une chose : finir tranquillement
La liste de nos rêves	2018-12-05	7.4	Calvin est un jeune bagagiste qui a abandonné ses études et qu
Toy Story 3	2010-06-16	7.8	Woody, Buzz l'Éclair et le reste des jouets reviennent dans un
Alice au pays des merveilles	2010-03-03	6.639	Alice, désormais âgée de 19 ans, retourne dans le monde fantasti
Ferdinand	2017-12-09	7.21	Ferdinand est un taureau au grand cœur. Victime de son imposant
L'Âge de glace : Les Lois de l'Univers	2016-06-23	6.1	La quête permanente de Scrat pour attraper son insaisissable no
Madagascar 3 : Bons baisers d'Europe	2012-06-06	6.612	Alex le lion, Marty le zèbre, Gloria l'hippopotame et Melman la

only showing top 20 rows

Nombre de votes	Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query
12534.0	162.582	Moana	en	Aventure, Comédie, Familial, Animation	NULL	NULL	NULL	NULL	NULL
2166.0	139.1	The Christmas Chronicles	en	Comédie, Aventure, Familial, Fantastique	NULL	NULL	NULL	NULL	NULL
19572.0	134.223	How to Train Your Dragon 2	en	Fantastique, Action, Aventure, Animation, Comédie, Familial	NULL	NULL	NULL	NULL	NULL
19042.0	149.827	Harry Potter and the Deathly Hallows: Part 1	en	Aventure, Fantastique	NULL	NULL	NULL	NULL	NULL
16582.0	134.957	Frozen	en	Animation, Familial, Aventure, Fantastique	NULL	NULL	NULL	NULL	NULL
13976.0	113.051	Pirates of the Caribbean: On Stranger Tides	en	Aventure, Action, Fantastique	NULL	NULL	NULL	NULL	NULL
36635.0	122.2	Inception	en	Action, Science-Fiction, Aventure	NULL	NULL	NULL	NULL	NULL
18770.0	97.924	WALL·E	en	Animation, Familial, Science-Fiction	NULL	NULL	NULL	NULL	NULL
6642.0	172.696	Rio	en	Animation, Aventure, Comédie, Familial	Mexico	2011	Movies	12	Rio
950.0	109.737	How to Train Your Dragon: Homecoming	en	Animation, Fantastique, Aventure, Action, Familial	NULL	NULL	NULL	NULL	NULL
11465.0	98.34	Tangled	en	Animation, Familial	NULL	NULL	NULL	NULL	NULL
5593.0	80.202	The Princess and the Frog	en	Animation, Romance, Fantastique, Familial	NULL	NULL	NULL	NULL	NULL
15655.0	102.64	Spider-Man: Into the Spider-Verse	en	Animation, Action, Aventure, Science-Fiction	NULL	NULL	NULL	NULL	NULL
20206.0	86.637	Up	en	Animation, Comédie, Familial, Aventure	NULL	NULL	NULL	NULL	NULL
476.0	79.63	Then Came You	en	Comédie, Drame	NULL	NULL	NULL	NULL	NULL
14647.0	75.883	Toy Story 3	en	Animation, Familial, Comédie	NULL	NULL	NULL	NULL	NULL
13929.0	83.225	Alice in Wonderland	en	Familial, Fantastique, Aventure	NULL	NULL	NULL	NULL	NULL
2836.0	113.396	Ferdinand	en	Animation, Familial, Aventure, Comédie	NULL	NULL	NULL	NULL	NULL
4531.0	67.368	Ice Age: Collision Course	en	Aventure, Animation, Familial, Comédie, Science-Fiction	NULL	NULL	NULL	NULL	NULL
6109.0	112.643	Madagascar 3: Europe's Most Wanted	en	Animation, Familial, Comédie, Aventure	NULL	NULL	NULL	NULL	NULL

Cette requête est similaire à une requête précédente, mais elle se concentre exclusivement sur les résumés des films. Elle permet d'identifier les films contenant certains mots-clés et de voir quelles catégories en ressortent. On remarque que les mots-clés extraits sont souvent associés aux genres aventure et animation. Cela

s'explique par le fait que des termes comme "magie" renvoient généralement à des univers fictifs, ce qui est plus courant dans les films d'animation ou les récits fantastiques où la liberté de création est plus grande.

De plus, cette requête a permis d'identifier le film Rio, qui figurait en deuxième position des recherches les plus effectuées au Mexique en 2011. On constate également que sa popularité est nettement plus élevée que celle des autres films identifiés, ce qui correspond bien au fait qu'il a été fortement recherché sur Google à cette période.

## E - Filtre final pour distinguer les films les plus recherchés et populaires

titre	Date de sortie	Note	Résumé
Interstellar	2014-11-05	8.4	Dans un futur proche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellaire pour franchir un trou de ver ;
Avengers : Infinity War	2018-04-25	8.2	Les Avengers et leurs alliés devront être prêts à tout sacrifier pour neutraliser le redoutable Thanos avant que son attaque éclair ne con
The Dark Knight : Le Chevalier noir	2008-07-16	8.516	Batman aborde une phase décisive de sa guerre contre le crime à Gotham City. Avec l'aide du lieutenant de police Jim Gordon et du nouveau ;
Joker	2019-10-01	8.1	Dans les années 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule dans la folie et devient le Joker.
Inception	2010-07-15	8.369	Dom Cobb est un voleur expérimenté, le meilleur dans l'art dangereux de l'extraction, voler les secrets les plus intimes enfouis au plus p
Coco	2017-10-27	8.2	Depuis déjà plusieurs générations, la musique est bannie dans la famille de Miguel. Un vrai déchirement pour le jeune garçon dont le rêve ;
Avengers : Endgame	2019-04-24	8.2	Après leur défaite face au Titan Thanos qui dans le film précédent s'est approprié toutes les pierres du Gant de l'infini, les Avengers et
Django Unchained	2012-12-25	8.2	Dans le sud des États-Unis, deux ans avant la guerre de Sécession, le Dr King Schultz, un chasseur de primes allemand, fait l'acquisition ;
Harry Potter et les Reliques de la mort - 2ème partie	2019-05-30	8.586	Toute la famille de Ki-taek est au chômage. Elle s'intéresse particulièrement au train de vie de la riche famille Park. Mais un incident
Le Loup de Wall Street	2011-07-12	8.092	Le combat entre les puissances du bien et du mal de l'univers des sorciers se transforme en guerre sans merci. L'école Poudlard est désorm
Vice-versa	2013-12-25	8.0	La licence de courtier en poche, et les narines déjà pleines de cocaïne, Jordan Belfort est prêt à conquérir Wall Street. Ce jour d'octobr
Deadpool	2015-06-17	7.9	Riley est déracinée du Mid-Ouest quand son père est tué à San Francisco. Comme tous, Riley est guidée par ses Emotions : Joie, Peur, Colèr
WALL-E	2016-02-09	7.624	Deadpool, est l'anti-héros le plus atypique de l'univers Marvel. À l'origine, il s'appelle Wade Wilson : un ancien militaire des Forces Sp
Spider-Man : New Generation	2008-06-22	8.1	Faites la connaissance de WALL-E (prononcez « Walli ») : WALL-E est le dernier être sur Terre et s'avère être un petit robot ! 700 ans plu
Gone Girl	2018-12-06	8.399	Spider-Man : New Generation suit les aventures de Miles Morales, un adolescent afro-américain et portoricain qui vit à Brooklyn et s'effor
Avengers	2014-10-01	7.892	Amy et Nick forment en apparence un couple modèle. Victimes de la crise, ils quittent Manhattan pour retourner s'installer dans la ville d
Harry Potter et les Reliques de la mort - 1ère partie	2012-04-25	7.723	Lorsque la sécurité et l'équilibre de la planète sont menacés par un ennemi d'un genre nouveau, Nick Fury, le directeur du SHIELD, l'agenci
Avatar	2016-08-26	8.485	Mitsuha est une lycéenne, la fille du maire d'une petite ville nichée entre les montagnes. Vivant avec sa petite sœur et sa grand-mère, c'es
	2010-11-17	7.747	Le pouvoir de Voldemort s'étend. Celui-ci contrôle maintenant le Ministère de la Magie et Poudlard. Harry, Ron et Hermione décident de terr
	2009-12-15	7.583	Un marine paralégique, envoyé sur la lune Pandora pour une mission unique, est tiraillé entre suivre ses ordres et protéger le monde qu'i

Popularité	original title	Original Language	Genres	Country	Year	Category	Rank	Query	Rank_Note	Rank_Votes	Rank_Popularity	Overall_Rank	row_num
238.105	Interstellar	en	Aventure, Drame, Science-Fiction	United States	2014	Movies	2	Interstellar	13	2	17	32	1
342.224	Avengers: Infinity War	en	Aventure, Action, Science-Fiction	NULL	NULL	NULL	NULL	NULL	46	30	3	79	1
142.229	The Dark Knight	en	Drame, Action, Crime, Thriller	NULL	NULL	NULL	NULL	NULL	7	10	97	114	1
201.229	Joker	en	Crime, Thriller, Drame	United States	2019	Movies	3	Joker	79	41	27	147	1
122.2	Inception	en	Action, Science-Fiction, Aventure	NULL	NULL	NULL	NULL	NULL	32	1	136	169	1
324.72	Coco	en	Familial, Animation, Musique, Aventure	NULL	NULL	NULL	NULL	NULL	46	134	4	184	1
124.211	Avengers: Endgame	en	Aventure, Science-Fiction, Action	France	2019	Les films	2	Avengers : Endgame	46	58	131	235	1
103.622	Django Unchained	en	Drame, Western	Switzerland	2013	Films	1	Django Unchained	46	34	178	258	1
10.555	Parasite	ko	Comédie, Thriller, Drame	NULL	NULL	NULL	NULL	NULL	8	150	152	310	1
143.398	Harry Potter and the Deathly Hallows: Part 2	en	Fantastique, Aventure	NULL	NULL	NULL	NULL	NULL	114	108	96	318	1
119.995	The Wolf of Wall Street	en	Crime, Drame, Comédie	NULL	NULL	NULL	NULL	NULL	133	60	138	331	1
126.962	Inside Out	en	Animation, Familial, Aventure, Drame, Comédie	NULL	NULL	NULL	NULL	NULL	180	88	129	397	1
198.413	Deadpool	en	Action, Aventure, Comédie	United Kingdom	2016	Movies	1	Deadpool	351	12	46	409	1
97.924	WALL-E	en	Animation, Familial, Science-Fiction	NULL	NULL	NULL	NULL	NULL	79	142	206	427	1
102.64	Spider-Man: Into the Spider-Verse	en	Animation, Action, Aventure, Science-Fiction	NULL	NULL	NULL	NULL	NULL	27	224	186	437	1
146.067	Gone Girl	en	Mystère, Thriller, Drame	United States	2014	Movies	4	Gone Girl	211	145	82	438	1
119.893	The Avengers	en	Science-Fiction, Action, Aventure	Poland	2012	Movies	4	Avengers	286	29	139	454	1
18.19	Your Name.	ja	Animation, Romance, Drame	NULL	NULL	NULL	NULL	NULL	10	343	142	495	1
149.827	Harry Potter and the Deathly Hallows: Part 1	en	Aventure, Fantastique	NULL	NULL	NULL	NULL	NULL	279	140	79	498	1
187.204	Avatar	en	Action, Aventure, Fantastique, Science-Fiction	NULL	NULL	NULL	NULL	NULL	426	11	63	500	1

Ce filtre de données nous semble être le plus pertinent pour analyser les films les plus populaires au fil des années. Nous avons d'abord appliqué un filtrage afin

d'obtenir des titres uniques, car sans cela, un film comme Interstellar apparaîtrait de nombreuses fois, étant donné qu'il a été populaire dans près d'une dizaine de pays.

Pour affiner cette analyse, nous avons mis en place un système de classement, ou ranking, en attribuant un rang à trois critères : la note moyenne, le nombre de votes et la popularité. Une quatrième colonne, `overall_rank`, a été créée en additionnant ces trois rangs. Plus cette valeur est faible, plus le film est considéré comme populaire. Par exemple, Interstellar obtient un `overall_rank` de 32, soit le plus bas de la base, ce qui en fait le film le plus populaire sur la période étudiée, de 2008 à 2020. Il s'agit d'un film d'aventure, de drame et de science-fiction, réalisé aux États-Unis et en langue anglaise, ce qui correspond aux tendances dominantes observées.

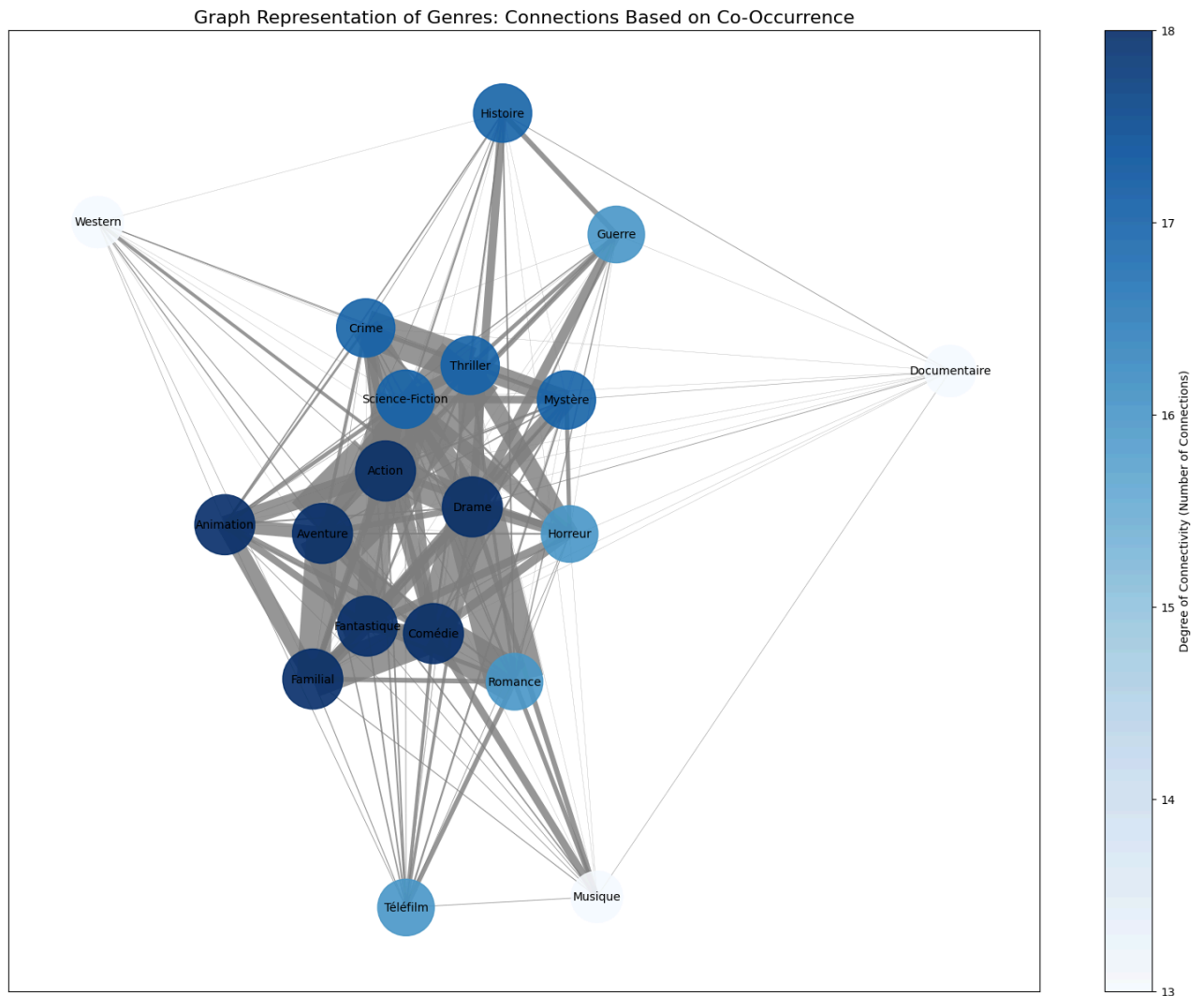
Cette méthode de ranking permet de réaliser des requêtes plus précises et d'éviter l'apparition de films faussement populaires, comme nous l'avons constaté précédemment avec des films ayant des notes très élevées mais obtenues sur un nombre de votes extrêmement faible.

Les tendances dégagées confirment la domination des films en langue anglaise, ainsi que la forte présence des genres animation, action, drame et science-fiction. Par ailleurs, plusieurs films apparaissant comme les plus recherchés sur Google figurent également parmi les films les plus populaires de notre classement, ce qui souligne la pertinence du croisement de ces deux bases de données.

Il serait possible d'améliorer cette analyse en intégrant davantage d'informations, notamment via un scraping direct des recherches Google, bien que nous ne disposions pas des ressources techniques nécessaires pour le mettre en place. De plus, malgré nos efforts pour harmoniser les titres et titres originaux afin d'optimiser la jointure des bases de données, il est possible que certaines correspondances n'aient pas été entièrement capturées. Cependant, les résultats obtenus jusqu'à présent restent cohérents et satisfaisants.

### **III - Graphs et table de fréquence sur nos données**





Pour comprendre ce graphe, plusieurs éléments doivent être pris en compte. Il représente les connexions entre différents genres associés à un même film, ce qui permet d'identifier les genres les plus liés dans notre base de données. Chaque couleur attribuée met en évidence ces connexions, avec une intensité plus marquée lorsque les genres sont fortement reliés entre eux.

L'épaisseur des arêtes représente le nombre de films dans lesquels deux genres apparaissent ensemble. Par exemple, on observe une forte connexion entre animation et aventure, tandis que la connexion entre aventure et musique est beaucoup plus faible. Afin d'éviter une surcharge d'informations, seuls les genres les plus connectés ont été affichés sur le graphe.

On distingue également des regroupements naturels entre certains genres. Par exemple, un groupe centré sur animation regroupe des genres comme famille et fantaisie, ce qui suggère que ces types de films sont souvent liés et difficiles à dissocier.

On remarque aussi que certains genres très présents, comme comédie et drame, sont fortement connectés à d'autres, car ils constituent des thèmes polyvalents dans l'industrie cinématographique. À l'inverse, les genres plus isolés ou faiblement connectés peuvent indiquer des catégories de films plus spécifiques ou de niche, comme musique.

Enfin, concernant les tendances des films les plus populaires, on observe que les genres ayant une forte corrélation reflètent certaines conventions dominantes dans l'industrie du cinéma. Par exemple, action est souvent lié à thriller, tandis que romance est fréquemment associé à drame. Ces connexions illustrent les préférences du public et les combinaisons de genres les plus exploitées par les studios.

## A - Les paires de genre les plus fréquentes

	Frequency	Genre 1	Genre 2
47	377	Action	Aventure
35	307	Action	Thriller
10	255	Action	Science-Fiction
24	255	Drame	Romance
0	228	Animation	Familial
25	218	Drame	Thriller
43	211	Aventure	Fantastique
1	195	Animation	Comédie
44	189	Horreur	Thriller
98	181	Comédie	Romance

Les données chiffrées sont plus faciles à interpréter et permettent de confirmer les observations faites à partir du graphe. On constate que le genre action est très souvent associé à d'autres genres, tout comme drame et animation. Ces catégories apparaissent comme des thèmes récurrents dans l'industrie du cinéma, illustrant les tendances dominantes et les combinaisons de genres les plus exploitées par les productions cinématographiques.

## B- Les 10 genres les plus connectés

	Genre	Degree
5	Action	2455
7	Drame	2373
8	Aventure	2074
2	Comédie	2014
9	Thriller	1851
3	Fantastique	1401
1	Familial	1333
0	Animation	1332
4	Science-Fiction	1231
6	Romance	972

Ici, nous pouvons observer les genres les plus connectés entre eux. Une fois de plus, on constate que le genre action est particulièrement lié à de nombreux autres genres, confirmant son omniprésence et son association fréquente avec différentes catégories cinématographiques.

## C - Filtre des co-occurrences avec action

	Frequency	Genre 1	Genre 2
47	377	Action	Aventure
35	307	Action	Thriller
10	255	Action	Science-Fiction
49	158	Action	Comédie
13	157	Action	Crime
69	140	Action	Drame
54	124	Action	Fantastique
8	96	Aventure	Action
6	95	Science-Fiction	Action
63	76	Animation	Action

Si nous examinons les genres les plus souvent associés à l'action, nous constatons que les plus fréquents sont l'aventure, le thriller et la science-fiction. Cependant, lorsque l'action est considérée comme un genre secondaire, elle est le plus souvent combinée avec l'aventure, la science-fiction et l'animation.

Il apparaît ainsi que la science-fiction et l'action sont presque inséparables, ces deux genres étant très souvent liés dans les productions cinématographiques. Cette forte corrélation témoigne des tendances du cinéma, où l'action est fréquemment utilisée pour renforcer l'intensité et le dynamisme des récits de science-fiction.

## Difficulté

La difficulté de ce projet résidait d'une part dans le choix du sujet et des bases de données les plus pertinentes, afin qu'elles s'enrichissent mutuellement et apportent une réelle valeur ajoutée, tout en évitant d'accumuler des bases de données redondantes.

Une autre difficulté a été l'extraction des données, notamment via le scraping, ainsi que la configuration de nos machines pour que les packages et bibliothèques nécessaires fonctionnent correctement. Certaines extractions et exécutions de code se sont révélées très coûteuses en temps et en ressources, ce qui a complexifié le traitement des données.

Nous avons également rencontré des limitations techniques liées à nos machines, qui ne permettaient pas d'importer directement certaines données, notamment avec

Visual Studio Code. Il a donc fallu adapter notre code et utiliser des alternatives, comme la conversion des données sous Pandas, afin de rendre le projet fonctionnel.

De plus, configurer airflow correctement nous a pris aussi énormément de temps dans l'avancement du projet.

Nous pensons qu'il serait possible d'améliorer ce projet en mettant en place une extraction automatique des informations issues des recherches Google, à l'image de ce que nous avons fait avec The Movie Database. Cela permettrait de capturer un volume de données plus important, d'affiner l'analyse et de rendre les résultats encore plus pertinents et représentatifs des tendances réelles.

## Conclusion

Ce projet nous a permis d'explorer en profondeur les tendances du cinéma en analysant deux bases de données complémentaires, The Movie Database et une base issue des recherches Google. L'objectif principal était d'étudier la popularité des films à travers leur notation, leur nombre de votes et leur présence dans les recherches effectuées sur Google afin de dégager des tendances pertinentes en fonction des périodes et des pays

Pour cela, nous avons fusionné ces deux bases de données après un travail de nettoyage et d'harmonisation des titres afin de garantir une correspondance optimale entre les films recensés et leurs recherches en ligne. Cette fusion nous a permis de disposer d'un jeu de données plus riche, capable de révéler des corrélations entre la popularité d'un film sur les plateformes spécialisées et l'intérêt du public mesuré par les recherches effectuées sur Internet

Nos analyses ont mis en évidence plusieurs tendances majeures. Tout d'abord, les films en langue anglaise dominent largement le marché, tandis que les productions en d'autres langues restent minoritaires. Certains genres comme l'action, le drame, l'animation et la science-fiction sont particulièrement populaires et souvent associés entre eux. Nous avons également observé que certains films comme Interstellar et Inception sont non seulement bien notés et très votés, mais aussi fortement recherchés sur Google, confirmant leur statut de films cultes à l'échelle mondiale

L'analyse des co-occurrences des genres a également permis d'identifier les liens forts entre certaines catégories comme action et thriller ou encore romance et drame qui sont souvent combinées dans les productions cinématographiques. De plus,

l'étude des tendances par pays a révélé des différences culturelles, notamment l'engouement pour les films de super-héros dans certaines régions du monde

Cependant, plusieurs défis ont été rencontrés tout au long du projet. Le choix et la fusion des bases de données ont nécessité un travail minutieux pour garantir leur complémentarité et éviter les doublons. L'extraction des données s'est révélée complexe, notamment en raison des limitations techniques et des performances des machines utilisées. Le traitement des données a également exigé de nombreuses optimisations, notamment pour réduire le temps d'exécution des requêtes et améliorer la pertinence des résultats obtenus

Une des principales limites de ce projet réside dans l'impossibilité d'extraire automatiquement les recherches effectuées sur Google en temps réel, ce qui aurait permis une analyse encore plus précise et dynamique des tendances cinématographiques. Une autre piste d'amélioration serait d'affiner encore davantage le nettoyage des données pour éviter certains biais liés aux différences de formats et d'encodages dans les bases fusionnées

Ce projet a démontré l'intérêt d'une approche basée sur les données pour mieux comprendre les dynamiques de popularité des films et des séries. Il a permis d'identifier des tendances majeures en fonction des années, des genres et des pays tout en mettant en évidence les forces et limites des bases de données utilisées. Avec des ressources supplémentaires et des techniques avancées d'extraction de données, il serait possible d'aller encore plus loin dans l'analyse, notamment en intégrant de nouvelles sources d'information ou en automatisant certaines tâches pour améliorer la qualité et la précision des résultats obtenus.

## Script

```

In [1]: import requests
import pandas as pd

api_key = 'b53d29a248e6ccdaefa5ae9b0f42f9b5' # Remplacez par votre clé A

# Étape 1 : Récupérer les genres de films
url_genres = f'https://api.themoviedb.org/3/genre/movie/list?api_key={api_key}'
response_genres = requests.get(url_genres)
data_genres = response_genres.json()

# Créer un dictionnaire des genres
genre_dict = {genre['id']: genre['name'] for genre in data_genres['genres']}

films = []

# Limite de pages à récupérer, car TMDB limite à 500 pages (vous pouvez aller plus loin si besoin)
for page in range(1, 501):
    url = f'https://api.themoviedb.org/3/discover/movie?api_key={api_key}&page={page}'
    response = requests.get(url)
    data = response.json()

    # Vérification et ajout des films dans la liste
    for movie in data['results']:
        # Convertir les genre_ids en noms
        genre_names = [genre_dict.get(genre_id, 'Unknown') for genre_id in movie['genre_ids']]

        films.append({
            'Titre': movie['title'],
            'Date de sortie': movie['release_date'],
            'Note': movie['vote_average'],
            'Résumé': movie['overview'],
            'Nombre de votes': movie['vote_count'],
            'Popularité': movie['popularity'],
            'Original Title': movie['original_title'],
            'Original Language': movie['original_language'],
            'Genres': ', '.join(genre_names), # Convertir la liste des genres en chaîne
        })

# Conversion en DataFrame et export en CSV
df = pd.DataFrame(films)
df.to_csv('10K_DMTB_with_genres.csv', index=False, encoding='utf-8')

print("Fichier CSV généré avec succès.")

```

Fichier CSV généré avec succès.

```

In [3]: !pip install pyspark

```

```

Requirement already satisfied: pyspark in /usr/local/lib/python3.11/dist-packages (3.5.4)
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.11/dist-packages (from pyspark) (0.10.9.7)

```

```
In [4]: from pyspark.sql import SparkSession

In [5]: spark = SparkSession.builder \
        .appName("PySpark on Colab") \
        .getOrCreate()

In [6]: # Charger les fichiers csv
spark_dixk = spark.read.csv("/content/drive/MyDrive/M2 MASTER/datalakes/1
spark_dixk.show(10)
spark_dixk.count()
```

```
+-----+-----+-----+-----+-----+
--+
```

	Titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	Original Title	Original Language	Genres
1113	Venom: The Last D...	2024-10-22	6.6	Eddie et Venom so...	12656.263	12656.263	Venom: The Last D...	en	Action, Science-F...
420	Vaiana 2	2024-11-21	6.864	Vaiana 2 embarque...	4485.032	4485.032	Moana 2	en	Animation, Aventu...
126	Elevation	2024-11-07	6.4	Les survivants d'...	2554.883	2554.883	Elevation	en	Action, Science-F...
120	Absolution	2024-10-31	6.1	Ancien boxeur, Th...	1846.723	1846.723	Absolution	en	Action, Crime, Th...
42	Watchmen: Chapter II	2024-11-25	7.738	Soupçonnant les é...	1439.232	1439.232	Watchmen: Chapter II	en	Animation, Mystèr...
54	Armor	2024-10-30	5.6	Un père et son fi...	1423.925	1423.925	Armor	en	Action, Crime, Th...
1006	Gladiator II	2024-11-05	6.7	Des années après ...	1262.621	1262.621	Gladiator II	en	Action, Aventure,...
58	Sidelined: The QB...	2024-11-29	6.6	NULL	1338.291	1338.291	Sidelined: The QB...	en	Comédie, Romance,...
31	Miraculous World ...	2024-11-14	8.71	Afin de sauver l'...	1219.844	1219.844	Miraculous World ...	fr	Animation, Aventu...
1201	Terrifier 3	2024-10-09	6.85	Après avoir survé...	922.731	922.731	Terrifier 3	en	Horreur, Thrille...

```
--+
```

only showing top 10 rows

Out[6]: 10084

```
In [7]: from pyspark.sql import SparkSession
from pyspark.sql.functions import when, col
# Échanger les valeurs des colonnes "titre" et "original title" selon la
spark_dixk = spark_dixk.withColumn(
    "temp",
    when(col("Original Language") != "en", col("titre")).otherwise(col("original title")).withColumn(
        "titre",
        when(col("Original Language") != "en", col("original title")).otherwise(col("titre")).withColumn(
            "original title",
            col("temp")
        ).drop("temp")

# Afficher le DataFrame après modification
print("DataFrame après échange :")
spark_dixk.show()
spark_dixk.count()
```

DataFrame après échange :

	titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genre
1113	Venom: The Last Dance	2024-10-22	6.6	Eddie et Venom so...	12656.263	12656.263	Venom: The Last Dance	en	Action, Science-F...
420	Vaiana 2	2024-11-21	6.864	Vaiana 2 embarque...	4485.032	4485.032	Moana 2	en	Animation, Aventure...
126	Elevation	2024-11-07	6.4	Les survivants d'...	2554.883	2554.883	Elevation	en	Action, Science-F...
120	Absolution	2024-10-31	6.1	Ancien boxeur, Th...	1846.723	1846.723	Absolution	en	Action, Crime, Thrille
42	Watchmen: Chapter II	2024-11-25	7.738	Soupçonnant les é...	1439.232	1439.232	Watchmen: Chapter II	en	Animation, Mystère...
54	Armor	2024-10-30	5.6	Un père et son fi...	1423.925	1423.925	Armor	en	Action, Crime, Thrille
1006	Gladiator II	2024-11-05	6.7	Des années après ...	1262.621	1262.621	Gladiator II	en	Action, Aventure, ...
58	Sidelined: The QB...	2024-11-29	6.6	NULL	1338.291	1338.291	Sidelined: The QB...	en	Comédie, Romance, ...
31	Miraculous World ...	2024-11-14	8.71	Afin de sauver l'...	1219.844	1219.844	Miraculous World ...	fr	Animation, Aventure...
1201	Terrifier 3	2024-10-09	6.85	Après avoir survé...	922.731	922.731	Terrifier 3	en	Horreur, Thrille



```

r|
|Apocalipsis Z: el...|    2024-10-04| 6.86|Lorsqu'une sorte ...|
677|    958.51|Apocalypse Z : le...|                es|Drame, Action, Ho...
|
|                Wicked|    2024-11-20|7.597|Elphaba, une jeun...|
485|    1084.787|                Wicked|                en|Drame, Romance, F...
|
|    Alien : Romulus|    2024-08-13| 7.2|Quelques années a...|
2588|    917.605|    Alien: Romulus|                en|Horreur, Science-...
.|
|Deadpool & Wolverine|    2024-07-24| 7.7|Un Wade Wilson dé...|
5839|    893.759|Deadpool & Wolverine|                en|Action, Comédie, ..
.|
|                Smile 2|    2024-10-16| 6.6|Sur le point d'en...|
888|    970.057|                Smile 2|                en|    Horreur, Mystère
|
|    The Substance|    2024-09-07|7.256|Elisabeth Sparkle...|
2475|    900.139|    The Substance|                en|Horreur, Science-...
.|
|                Red One|    2024-10-31| 6.6|Après l'enlèvemen...|
227|    864.206|                Red One|                en|Action, Comédie, ...
|
|    Vice-versa 2|    2024-06-11| 7.6|Riley, adolescent...|
4935|    740.187|    Inside Out 2|                en|Animation, Aventu...
.|
|                Levels|    2024-11-01|5.689|Après avoir été t...|
37|    792.543|                Levels|                en|Action, Science-F...|
|Transformers : Le...|    2024-09-11| 8.1|Les origines d'Op...|
843|    725.848|    Transformers One|                en|Animation, Scienc...
|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
--+
only showing top 20 rows

```

Out[7]: 10084

```

In [8]: spark_google = spark.read.csv("/content/drive/MyDrive/M2 MASTER/datalakes
spark_google.show()
spark_google.count()

```

Country	Year	Category	Rank	Query
Global	2001	Movies	1	Harry Potter
Global	2001	Movies	2	Lord of the Rings
Global	2001	Movies	3	Final Fantasy
Global	2001	Movies	4	Tomb Raider
Global	2001	Movies	5	Shrek
Global	2002	Movies	1	Spiderman
Global	2002	Movies	2	Harry Potter
Global	2002	Movies	3	Star Wars
Global	2002	Movies	4	Jackass
Global	2002	Movies	5	Scooby Doo
Australia	2003	Popular Movies	1	2 Fast 2 Furious
Australia	2003	Popular Movies	2	Lord of the Rings
Australia	2003	Popular Movies	3	Finding Nemo
Australia	2003	Popular Movies	4	Texas Chainsaw Ma...
Australia	2003	Popular Movies	5	Pirates of the Ca...
Germany	2003	Beliebte Filme	1	Matrix
Germany	2003	Beliebte Filme	2	Harry Potter
Germany	2003	Beliebte Filme	3	Herr Der Ringe
Germany	2003	Beliebte Filme	4	Findet Nemo
Germany	2003	Beliebte Filme	5	Kill Bill

only showing top 20 rows

Out[8]: 880

In [7]:

```
In [9]: from pyspark.sql import SparkSession
# Joindre les DataFrames sur les colonnes "original title" et "query"
# La jointure se fait entre "original title" dans spark_10k et "query" dans spark_google
join_condition = spark_dixk["Titre"] == spark_google["query"]

# Effectuer la jointure (ici, on fait une jointure gauche)
joined_df = spark_dixk.join(spark_google, join_condition, "left")

# Créer le DataFrame final 10K_google
dixK_google = joined_df

# Afficher les résultats de la jointure
dixK_google.show(10)

dixK_google.count()
```

```

+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|          titre|Date de sortie| Note|          Résumé|Nombre de
votes|Popularité|          original title|Original Language|          Genr
es|Country|Year|Category|Rank|Query|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
|Venom: The Last D...| 2024-10-22| 6.6|Eddie et Venom so...|
1113| 12656.263|Venom: The Last D...|          en|Action, Science-F..
.| NULL|NULL| NULL|NULL| NULL|
|          Vaiana 2| 2024-11-21|6.864|Vaiana 2 embarque...|
420| 4485.032|          Moana 2|          en|Animation, Aventu...
| NULL|NULL| NULL|NULL| NULL|
|          Elevation| 2024-11-07| 6.4|Les survivants d'...|
126| 2554.883|          Elevation|          en|Action, Science-F...
| NULL|NULL| NULL|NULL| NULL|
|          Absolution| 2024-10-31| 6.1|Ancien boxeur, Th...|
120| 1846.723|          Absolution|          en|Action, Crime, Th...
| NULL|NULL| NULL|NULL| NULL|
|Watchmen: Chapter II| 2024-11-25|7.738|Soupçonnant les é...|
42| 1439.232|Watchmen: Chapter II|          en|Animation, Mystèr...|
NULL|NULL| NULL|NULL| NULL|
|          Armor| 2024-10-30| 5.6|Un père et son fi...|
54| 1423.925|          Armor|          en|Action, Crime, Th...|
NULL|NULL| NULL|NULL| NULL|
|          Gladiator II| 2024-11-05| 6.7|Des années après ...|
1006| 1262.621|          Gladiator II|          en|Action, Aventure,...
.| NULL|NULL| NULL|NULL| NULL|
|Sidelined: The QB...| 2024-11-29| 6.6|          NULL|
58| 1338.291|Sidelined: The QB...|          en|Comédie, Romance,...|
NULL|NULL| NULL|NULL| NULL|
|Miraculous World ...| 2024-11-14| 8.71|Afin de sauver l'...|
31| 1219.844|Miraculous World ...|          fr|Animation, Aventu...|
NULL|NULL| NULL|NULL| NULL|
|          Terrifier 3| 2024-10-09| 6.85|Après avoir survé...|
1201| 922.731|          Terrifier 3|          en|          Horreur, Thrille
r| NULL|NULL| NULL|NULL| NULL|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
only showing top 10 rows

```

Out[9]: 10341

```

In [10]: from pyspark.sql import SparkSession
from pyspark.sql.functions import year, col

spark_dixk_filtered = dixK_google.filter(
    (year(col("Date de sortie")) >= 2008) &
    (year(col("Date de sortie")) <= 2020)
)

```

```
# Afficher le DataFrame filtré
print("DataFrame filtré (années 2008-2020) :")
spark_dixk_filtered.show(20)

spark_dixk_filtered.count()
```

DataFrame filtré (années 2008-2020) :

	titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité	original title	Original Language	Genre
	Country	Year	Category	Rank	Query				
	Le Grinch	2018-11-08	6.864	Terré dans sa cav...					
3944	477.459	The Grinch		en	Animation, Famili..				
..	NULL	NULL		NULL	NULL				
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	United Kingdom	2018		Movies	5	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Singapore	2018		Top Trending Movies	5	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Romania	2018		Filme	3	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Nigeria	2018		Movies	4	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Malaysia	2018		Filem	3	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Hungary	2018		Filmek	1	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	France	2018		Les films	5	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Canada	2018		Movies	4	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Brazil	2018		Filmes	5	Venom			
	Venom	2018-09-28	6.8	Des symbiotes déb...					
16065	301.84	Venom		en	Science-Fiction, .				
..	Global	2018		Movies	3	Venom			
	일본 엄마	2017-02-09	6.2			NULL			
28	364.704	일본 엄마		ko	Romance, Drame				
	NULL	NULL		NULL					
	Des jours plus be...	2010-02-19	6.523	Wim et Floor, un ...					

```

22| 299.557|Des jours plus be...| nl| Romance|
NULL|NULL| NULL|NULL| NULL|
|Avengers : Infini...| 2018-04-25| 8.2|Les Avengers et l...|
29791| 342.224|Avengers: Infinit...| en|Aventure, Action,.
..| NULL|NULL| NULL|NULL| NULL|
| Equalizer| 2014-09-24|7.281|McCall, un homme ...|
9029| 192.911| The Equalizer| en|Thriller, Action,..
.| NULL|NULL| NULL|NULL| NULL|
| Interstellar| 2014-11-05| 8.4|Dans un futur pro...|
35760| 238.105| Interstellar| en|Aventure, Drame, .
..| United States|2014| Movies| 2|Interstellar|
| Interstellar| 2014-11-05| 8.4|Dans un futur pro...|
35760| 238.105| Interstellar| en|Aventure, Drame, .
..|United Arab Emirates|2014| Movies| 2|Interstellar|
| Interstellar| 2014-11-05| 8.4|Dans un futur pro...|
35760| 238.105| Interstellar| en|Aventure, Drame, .
..| Sweden|2014| Film| 1|Interstellar|
| Interstellar| 2014-11-05| 8.4|Dans un futur pro...|
35760| 238.105| Interstellar| en|Aventure, Drame, .
..| Singapore|2014| Movies| 1|Interstellar|
| Interstellar| 2014-11-05| 8.4|Dans un futur pro...|
35760| 238.105| Interstellar| en|Aventure, Drame, .
..| Malaysia|2014| Movies (Filem)| 3|Interstellar|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
only showing top 20 rows

```

Out[10]: 4077

# 19/01/2025

## Requêtes simple

### les films les mieux notés

```

In [10]: from pyspark.sql.window import Window
from pyspark.sql.functions import row_number, col
from pyspark.sql.functions import max
from pyspark.sql.types import FloatType

spark_dixk_filtered = spark_dixk_filtered.withColumn("Popularité", col("P
.withColumn("Nombre de votes", c

In [11]: top_movies = spark_dixk_filtered.orderBy(col("Note").desc()).limit(10)
top_movies.show()

```

## ## film les plus populaires

titre	Date de sortie	Note	Résumé
original title			Original
Number of votes	Popularity		

Language	Genres		Country	Year	Category	Rank	Query
+-----+-----+-----+-----+-----+-----+-----+-----+							
+-----+-----+-----+-----+-----+-----+-----+-----+							
	Wedding Nightmare		2019-08-21	7.0	La nuit de noces ...		
4531	99.745	Ready or Not			en		
Horreur, Comédie	NULL NULL	NULL NULL	NULL				
	Ça		2017-09-06	7.2	À Derry, dans le ...		
19038	99.642	It			en		
Horreur	NULL NULL	NULL NULL	NULL				
	Ant-Man et la Guêpe		2018-07-04	6.93	Après les événeme...		
13263	99.172	Ant-Man and the Wasp			en		
Action, Aventure,...	NULL NULL	NULL NULL	NULL				
	부산행		2016-07-20	7.757	Tandis qu'un viru...		
7438	98.662	Dernier train pou...			ko H		
orreur, Thriller...	NULL NULL	NULL NULL	NULL				
	Le Flingueur		2011-01-13	6.523	Un tueur à gages ...		
3233	98.358	The Mechanic			en A		
ction, Thriller,...	NULL NULL	NULL NULL	NULL				
	Raiponce		2010-11-24	7.6	Quand le bandit l...		
11465	98.34	Tangled			en		
Animation, Familial	NULL NULL	NULL NULL	NULL				
	Geostorm		2017-10-12	6.1	Le jour où des sa...		
4774	98.239	Geostorm			en A		
ction, Science-F...	NULL NULL	NULL NULL	NULL				
劇場版「鬼滅の刃」無限列車編	2020-10-16	8.211	Le groupe de Tanj...				
3951	98.036	Demon Slayer : Ki...			ja A		
animation, Action...	NULL NULL	NULL NULL	NULL				
	Taken 3		2014-12-16	6.3	L'ex-agent spécia...		
5735	98.015	Taken 3			en		
Thriller, Action	NULL NULL	NULL NULL	NULL				
	WALL·E		2008-06-22	8.1	Faites la connais...		
18770	97.924	WALL·E			en		
Animation, Famili...	NULL NULL	NULL NULL	NULL				
	Homefront		2013-11-12	6.944	Un agent de la DE...		
3484	97.91	Homefront			en		
Action, Thriller	NULL NULL	NULL NULL	NULL				
	Spider-Man : Home...		2017-07-05	7.3	Après ses spectac...		
21776	97.905	Spider-Man: Homec...			en		
Action, Aventure,...	NULL NULL	NULL NULL	NULL				
	엄마친구		2015-08-20	6.4	NULL		
16	97.245	엄마친구			ko		
Drame	NULL NULL	NULL NULL	NULL				
	Thor : Ragnarok		2017-10-02	7.6	Thor est emprison...		
20722	97.189	Thor: Ragnarok			en		
Action, Aventure,...	NULL NULL	NULL NULL	NULL				
	John Wick 2		2017-02-08	7.335	John Wick est for...		
13025	96.415	John Wick: Chapter 2			en		
Action, Thriller,...	NULL NULL	NULL NULL	NULL				
	Taken 2		2012-09-27	6.357	L'agent de la CIA...		
6620	96.391	Taken 2			en A		
ction, Crime, Th...	NULL NULL	NULL NULL	NULL				
	劇場版 NARUTO -ナルト- ...		2009-08-01	7.319	Parmi les cinq gr...		

```

459|    96.099|Naruto Shippuden ...|          ja|Ac
tion, Aventure,...|    NULL|NULL|    NULL|NULL| NULL|
|          Kiss and Kill|    2017-08-08|5.473|          NULL|
56|    96.027|    Kiss and Kill|          en|Mys
tère, Thriller...|    NULL|NULL|    NULL|NULL| NULL|
|          Legend|    2015-09-09|7.059|Londres, les anné...|
3854|    96.006|          Legend|          en|
Crime, Thriller|    NULL|NULL|    NULL|NULL| NULL|
|          20.252|    2008-08-19|    6.7|"1969 : Kenji et ...| l
a fin du monde ...|    96|    20th Century Boys|20世紀少年 -第1章- 終わ
りの始まり|          ja|    NULL|NULL|    NULL|NULL| NULL|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+

```

## film avec le plus de votes

```

In [13]: nbr_votes_movies = spark_dixk_filtered.orderBy(col("Nombre de votes").des
nbr_votes_movies.show()

```



## popularité moyenne par année

Page 12 sur 52

```
# Now calculate popularity by year
popularity_by_year = spark_dixk_filtered_popu.groupBy(year(col("Date de s
popularity_by_year.show()
```

```
+-----+-----+
|Year|    avg(Popularité)|
+-----+-----+
|2008|47.482321126837476|
|2009|38.556854092512005|
|2010|123.13161097809859|
|2011| 43.98138288979058|
|2012| 37.12952691252514|
|2013| 35.5063122799522|
|2014| 52.13135466752229|
|2015|60.270837715861255|
|2016|43.484417385243354|
|2017|38.310180148807184|
|2018| 49.97792479261082|
|2019| 47.21958457897236|
|2020| 37.49174136457772|
+-----+-----+
```

## compter le nombre de film par genre

In [56]: `from pyspark.sql.functions import col, split, explode, array_distinct, mo`

```
# Explorer les genres en colonnes multiples
genre_counts = spark_dixk_filtered.withColumn("Genre", explode(split(col(
    .groupBy("Genre").count().orderBy(col("count").desc())
genre_counts.show()
```

Genre	count
Drame	787
Thriller	771
Drame	752
Action	688
Aventure	640
Comédie	605
Comédie	495
Action	461
Romance	444
Fantastique	444
Science-Fiction	417
Familial	378
Horreur	312
Crime	310
Animation	302
Mystère	267
Horreur	239
Thriller	229
Romance	206
Animation	197

only showing top 20 rows

## film d'un pays spécifique

```
In [47]: movies_by_country = spark_dixk_filtered.filter(col("Country") == "United
movies_by_country.show()
```

titre	Date de sortie	Note	Résumé	Nombre de votes
Popularité	original title	Original Language	Genres	
Country	Year	Category	Rank	Query
Venom	2018-09-28	6.8	Des symbiotes déb...	16065
301.84	Venom	en	Science-Fiction, ...	United
Kingdom 2018	Movies	5	Venom	
Joker	2019-10-01	8.1	Dans les années 1...	25781
201.229	Joker	en	Crime, Thriller, ...	United
Kingdom 2019	Movies	2	Joker	
Deadpool	2016-02-09	7.624	Deadpool, est l'a...	31055
190.413	Deadpool	en	Action, Aventure, ...	United
Kingdom 2016	Movies	1	Deadpool	
American Sniper	2014-12-25	7.426	Tireur d'élite de...	12815
168.805	American Sniper	en	Guerre, Action	United

```

Kingdom|2015| Films| 5|American Sniper|
| Black Panther| 2018-02-13|7.382|Après les événeme...| 22164
| 138.884| Black Panther| en|Action, Aventure,...|United
Kingdom|2018| Movies| 1| Black Panther|
| Deadpool 2| 2018-05-10|7.492|Deadpool se voit ...| 17954
| 143.977| Deadpool 2| en|Action, Comédie, ...|United
Kingdom|2018| Movies| 4| Deadpool 2|
| World War Z| 2013-06-19|6.818|Gerry Lane se lan...| 15342
| 81.861| World War Z| en|Action, Horreur, ...|United
Kingdom|2013| Movies| 2| World War Z|
| The Revenant| 2015-12-25|7.533|Dans une Amérique...| 18053
| 106.682| The Revenant| en|Western, Drame, A...|United
Kingdom|2016| Movies| 3| The Revenant|
| Jurassic World| 2015-06-06|6.692|Après le décès de...| 20315
| 69.709| Jurassic World| en|Action, Aventure,...|United
Kingdom|2015| Films| 1| Jurassic World|
| Iron Man 3| 2013-04-18|6.931|Tony Stark, l'ind...| 22142
| 89.501| Iron Man 3| en|Action, Aventure,...|United
Kingdom|2013| Movies| 3| Iron Man 3|
| Suicide Squad| 2016-08-03| 5.9|C'est tellement j...| 21093
| 75.705| Suicide Squad| en|Action, Aventure,...|United
Kingdom|2016| Movies| 2| Suicide Squad|
| Spectre| 2015-10-26| 6.5|Un message crypti...| 10480
| 68.839| Spectre| en|Action, Aventure,...|United
Kingdom|2015| Films| 4| Spectre|
| Man of Steel| 2013-06-12| 6.6|Un petit garçon d...| 15094
| 45.167| Man of Steel| en|Action, Aventure,...|United
Kingdom|2013| Movies| 1| Man of Steel|
| Captain Marvel| 2019-03-06| 6.8|Captain Marvel ra...| 15685
| 58.773| Captain Marvel| en|Action, Aventure,...|United
Kingdom|2019| Movies| 3| Captain Marvel|
| Joker| 2015-01-14|5.668|Nick Wild, ex-mar...| 1768
| 40.869| Wild Card| en|Thriller, Crime, ...|United
Kingdom|2019| Movies| 2| Joker|
| 1917| 2019-12-25|7.991|Pris dans la tour...| 12378
| 35.96| 1917| en|Guerre, Histoire,...|United
Kingdom|2020| Movies| 1| 1917|
| The Gentlemen| 2020-01-01| 7.68|Quand Mickey Pear...| 5824
| 37.536| The Gentlemen| en|Action, Crime, Co...|United
Kingdom|2020| Movies| 3| The Gentlemen|
| Jojo Rabbit| 2019-10-18|8.017|Jojo est un petit...| 9527
| 31.174| Jojo Rabbit| en|Comédie, Guerre, ...|United
Kingdom|2020| Movies| 5| Jojo Rabbit|
| True Grit| 2010-12-22|7.306|1870, juste après...| 5232
| 24.246| True Grit| en|Drame, Aventure, ...|United
Kingdom|2011| Movies| 4| True Grit|
| Lion| 2016-11-24|8.068|Une incroyable hi...| 6533
| 25.71| Lion| en| Drame|United
Kingdom|2020| Movies| 4| Lion|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+

```

## film contenant un mot clé dans le titre ou le résumé

```
In [17]: keyword_movies = spark_dixk_filtered.filter(
          (col("titre").contains("Venom")) | (col("Résumé").contains("symbiotes")
        )
        keyword_movies.show()
```

titre	Date de sortie	Note	Résumé	Nombre de votes	Popularité
original title	Original Language		Genres		Country
Category	Rank	Query			Year
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		United Kingdom 2018
Movies	5	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Singapore 2018
Top Trending Movies	5	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Romania 2018
Filme	3	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Nigeria 2018
Movies	4	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Malaysia 2018
Filem	3	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Hungary 2018
Filmek	1	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		France 2018
Les films	5	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Canada 2018
Movies	4	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Brazil 2018
Filmes	5	Venom			
Venom	2018-09-28	6.8	Des symbiotes déb...	16065.0	301.84
Venom			en Science-Fiction, ...		Global 2018
Movies	3	Venom			

2. Le top des films par note moyenne, pour chaque année entre 2008 et 2020

[illegible]

```

-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
|The Dark Knight : Le Chevalier noir|2008-07-16
|8.516|Batman aborde une phase décisive de sa guerre contre le crime à Got
ham City. Avec l'aide du lieutenant de police Jim Gordon et du nouveau pro
cureur Harvey Dent, il entreprend de démanteler les dernières organisation
s criminelles qui infestent les rues de la ville. L'association s'avère ef
ficace, mais le trio se heurte bientôt à un nouveau génie du crime qui rép
and la terreur et le chaos dans Gotham : le Joker. On ne sait pas d'où il
vient ni qui il est. Ce criminel possède une intelligence redoutable doubl
é d'un humour sordide et n'hésite pas à s'attaquer à la pègre locale dans
le seul but de semer le chaos.
|32971.0|142.229|The Dark Knight
|en|Drame, Action, Crime, Thriller|NULL|NULL|NU
LL|1|NULL|
|Inglourious Basterds|2009-08-02
|8.218|Dans la France occupée de 1940, un groupe de soldats juifs américai
ns, les bâtards, mène des actions punitives particulièrement sanglantes co
ntre les nazis du Troisième Reich. Leur chemin rencontre celui de Shosanna
Dreyfus, une jeune juive réfugiée à Paris exploitante d'une salle de ciném
a qui entend elle aussi assouvir sa propre vengeance envers les nazis.
|22250.0|66.16|Inglourious Basterds
|en|Drame, Thriller, Guerre|NULL|NULL|NU
LL|1|NULL|
|Inception|2010-07-15
|8.369|Dom Cobb est un voleur expérimenté, le meilleur dans l'art dangereu
x de l'extraction, voler les secrets les plus intimes enfouis au plus prof
ond du subconscient durant une phase de rêve, lorsque l'esprit est le plus
vulnérable. Les capacités de Cobb ont fait des envieux dans le monde tourm
enté de l'espionnage industriel alors qu'il devient fugitif en perdant tou
t ce qu'il a un jour aimé. Une chance de se racheter lui est alors offerte
. Une ultime mission grâce à laquelle il pourrait retrouver sa vie passée
mais uniquement s'il parvient à accomplir l'impossible inception.
|36635.0|122.2|Inception
|en|Action, Science-Fiction, Aventure|NULL|NULL|NU
LL|1|NULL|
|Intouchables|2011-11-02
|8.275|À la suite d'un accident de parapente, Philippe, riche aristocrate,
engage comme aide à domicile Driss, un jeune de banlieue tout juste sorti
de prison. Bref la personne la moins adaptée pour le job. Ensemble ils von
t faire cohabiter Vivaldi et Earth Wind and Fire, le verbe et la vanne, le
s costumes et les bas de survêtement... Deux univers vont se télescoper, s'a
pprivoiser, pour donner naissance à une amitié aussi dingue, drôle et fort
e qu'inattendue, une relation unique qui fera des étincelles et qui les re
ndra... Intouchables.
|17227.0|74.526|Intouchables
|fr|Drame, Comédie|NULL|NULL|NU
LL|1|NULL|
|Django Unchained|2012-12-25

```

|8.2 |Dans le sud des États-Unis, deux ans avant la guerre de Sécession, le Dr King Schultz, un chasseur de primes allemand, fait l'acquisition de Django, un esclave qui peut l'aider à traquer les frères Brittle, les meurtriers qu'il recherche. Schultz promet à Django de lui rendre sa liberté lorsqu'il aura capturé les Brittle – morts ou vifs. Alors que les deux hommes pistent les dangereux criminels, Django n'oublie pas que son seul but est de retrouver Broomhilda, sa femme, dont il fut séparé à cause du commerce des esclaves... Lorsque Django et Schultz arrivent dans l'immense plantation du puissant Calvin Candie, ils éveillent les soupçons de Stephen, un esclave qui sert Candie et a toute sa confiance. Le moindre de leurs mouvements est désormais épié par une dangereuse organisation de plus en plus proche... Si Django et Schultz veulent espérer s'enfuir avec Broomhilda, ils vont devoir choisir entre l'indépendance et la solidarité, entre le sacrifice et la survie...|26237.0 |103.622 |Django Unchained

|en |Drame, Western |Switzerland |2013|Film  
|1 |Django Unchained|

|No Half Measures: Creating the Final Season of Breaking Bad|2013-11-26

|8.4 |Un documentaire sur la réalisation de la cinquième saison de la célèbre série du diffuseur télévisé AMC, Breaking Bad.

|111.0 |20.413 |No Half Measures: Creating the Final Season of Breaking Bad|en |Documentaire |NULL

|NULL|NULL |1 |NULL |

|Interstellar |2014-11-05

|8.4 |Dans un futur proche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellaire pour franchir un trou de ver permettant de parcourir des distances jusque-là infranchissables. Leur but : trouver un nouveau foyer pour l'humanité.

|35760.0 |238.105 |Interstellar

|en |Aventure, Drame, Science-Fiction |United States|2014|Movies |1 |Interstellar |

|ハイキュー!! 終わりと始まり |2015-07-03 |8.

6 |Premier Haikyuu !! film récapitulatif. Ld

|36.0 |13.696 |Haikyuu !! – Film 1 – Un début et une fin

|ja |Animation, Drame |NULL |NULL|NULL  
LL |1 |NULL |

|働く女の艶めかしい完全着衣ファック 水野朝陽 |2016-12-13 |9.5

|NULL

|2.0 |38.398 |働く女の艶めかしい完全着衣ファック 水野朝陽

|ja |Drame |NULL |NULL|NULL  
LL |1 |NULL |

|Sree Hari Hara Sudhan Swami Ayyappan |2017-04-17

|9.0 |NULL

|3.0 |38.932 |Sree Hari Hara Sudhan Swami Ayyappan

|ta |Animation, Drame, Histoire |NULL |NULL|NULL  
LL |1 |NULL |

|ネトラセラレ3 ～愛～ |2018-04-02 |9

.0 |NULL

|1.0 |37.372 |ネトラセラレ3 ～愛～

|ja |Drame |NULL |NULL|NULL  
LL |1 |NULL |

|여자 하숙집 3 |2019-06-11 |

9.8 |NULL

|2.0 |21.277 |여자 하숙집 3



[illegible]

3. Analyser la corrélation entre la popularité et le nombre de votes
4. Les 3 pays les plus prolifiques par nombre de films

```
In [57]: # Compter les films par pays
country_movie_count = spark_dikk_filtered.groupBy("Country") \
    .count() \
    .orderBy(col("count").desc())

# Afficher les 3 pays les plus prolifiques
country_movie_count.show(10, truncate=False)
```

Country	count
NULL	3756
Global	29
France	27
Singapore	21
Ireland	21
United Kingdom	20
United States	18
Malaysia	15
Canada	14
Sweden	13

only showing top 10 rows

#### 4.5 Les 3 pays les plus prolifiques par original language

```
In [60]: # Compter les films par pays
country_movie_count = spark_dixk_filtered.groupBy("Original Language") \
    .count() \
    .orderBy(col("count").desc())

# Afficher les 3 pays les plus prolifiques
country_movie_count.show(20, truncate=False)
```

Original Language	count
en	3008
ko	207
ja	201
fr	134
es	96
zh	45
de	33
NULL	29
hi	25
ru	23
it	17
nl	16
cn	14
da	14
pt	10
th	9
no	9
tl	8
id	8
sv	7

only showing top 20 rows

## 5. Les films les mieux notés par catégorie

```
In [65]: # Trouver les films avec les meilleures notes pour chaque catégorie
category_top_movies = spark_dixk_filtered.withColumn("Rank", row_number()
    .filter(col("Rank") == 1)

category_top_movies.show(truncate=False)
```

titre	Date de sortie	Note	Résumé		
Nombre de votes	Popularité	original title	Original Language	Genres	
Country	Year	Category	Rank	Query	ID
여자 하숙집 3	2019-06-11	9.8	NULL		
2	21.277	여자 하숙집 3	ko	Romance	
NULL	NULL	NULL	1	NULL	2938
Skyfall	2012-10-24	7.2	Lorsque la dernière mission de Bond tourne mal, plusieurs agents infiltrés se retrouvent exposés dans le monde entier. Le MI6 est attaqué, et M est obligée de relocaliser l'Agence. Ces événements ébranlent son autorité, et elle est remise en cause par Malory, le nouveau président de l'ISC, le comité chargé du renseignement et de la sécurité. Le MI6 est à présent sous le coup d'une double menace, intérieure et extérieure. Il ne reste à M qu'un seul allié de confiance vers		

qui se tourner: Bond. Plus que jamais, 007 va devoir agir dans l'ombre. Avec l'aide d'Ève, un agent de terrain, il se lance sur la piste du mystérieux Silva, dont il doit identifier coûte que coûte l'objectif secret et mortel.

15160	105.629	Skyfall	en	Action, Av
enture, Thriller	Global	2012	Feature Films	1  Skyfa
ll	315			

|Joker |2019-10-01 |8.1 |Dans les années 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule dans la folie et devient le Joker.

25781	201.229	Joker	en	Crime, Thr
iller, Drame	Malaysia	2019	Filem	1  Joker
58				

|Wonder Woman |2017-05-30 |7.2 |Avant d'être Wonder Woman, elle s'appelait Diana, princesse des Amazones, entraînée pour être une guerrière impossible à conquérir. Elle est élevée sur une île isolée et paradisiaque, mais lorsqu'un pilote américain s'écrase sur leur rivage et annonce qu'un conflit à grande échelle fait rage dans le monde, Diana quitte son foyer, convaincue qu'elle doit arrêter cette menace. Combattant aux côtés de cet homme et des siens pour mettre fin à cette guerre et à toutes les guerres, Diana découvre ses vrais pouvoirs... Et son véritable destin.

19895	81.74	Wonder Woman	en	Action, Av
enture, Fantastique	Malaysia	2017	Filem (Movies)	1  Wonde
r Woman	474			

|Interstellar |2014-11-05 |8.4 |Dans un futur proche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellaire pour franchir un trou de ver permettant de parcourir des distances jusque-là infranchissables. Leur but : trouver un nouveau foyer pour l'humanité.

35760	238.105	Interstellar	en	Aventure,
Drame, Science-Fiction	Sweden	2014	Film	1  Inter
stellar	17			

|Joker |2019-10-01 |8.1 |Dans les années 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule dans la folie et devient le Joker.

25781	201.229	Joker	en	Crime, Thr
iller, Drame	Lithuania	2019	Filmai	1  Joker
59				

|Django Unchained |2012-12-25 |8.2 |Dans le sud des États-Unis, deux ans avant la guerre de Sécession, le Dr King Schultz, un chasseur de primes allemand, fait l'acquisition de Django, un esclave qui peut l'aider à traquer les frères Brittle, les meurtriers qu'il recherche. Schultz promet à Django de lui rendre sa liberté lorsqu'il aura capturé les Brittle – morts ou vifs. Alors que les deux hommes pistent les dangereux criminels, Django n'oublie pas que son seul but est de retrouver Broomhilda, sa femme, dont il fut séparé à cause du commerce des esclaves... Lorsque Django et Schultz arrivent dans l'immense plantation du puissant Calvin Candie, ils éveillent les soupçons de Stephen, un esclave qui sert Candie et a toute sa confiance. Le moindre de leurs mouvements est désormais épié par une dangereuse organisation de plus en plus proche... Si Django et Schultz veulent espérer s'enfuir avec Broomhilda, ils vont devoir choisir entre l'indépendance et la solidarité, entre le sacrifice et la survie...|26237 |103.622

Django Unchained	en	Drame, Western	Swi
------------------	----	----------------	-----

tzerland |2013|Filme |1 |Django Unchained |316 |  
 |Venom |2018-09-28 |6.8 |Des symbiotes débarquent sur la T  
 erre, parmi eux, Venom, qui va s'allier avec Eddie Brock. De son côté, un  
 puissant scientifique tente de faire évoluer l'humanité avec les symbiotes  
 , le duo d'anti-héros va devoir tout faire pour l'arrêter !  
 |16065 |301.84 |Venom |en |Science-Fi  
 ction, Action |Hungary |2018|Filmek |1 |Venom  
 |6 |  
 |Deadpool |2016-02-09 |7.624|Deadpool, est l'anti-héros le plu  
 s atypique de l'univers Marvel. À l'origine, il s'appelle Wade Wilson : un  
 ancien militaire des Forces Spéciales devenu mercenaire. Après avoir subi  
 une expérimentation hors norme qui va accélérer ses pouvoirs de guérison,  
 il va devenir Deadpool. Armé de ses nouvelles capacités et d'un humour noi  
 r survolté, Deadpool va traquer l'homme qui a bien failli anéantir sa vie.  
 |31055 |190.413 |Deadpool |en |Action, Av  
 enture, Comédie |Sweden |2016|Filmer |1 |Deadp  
 ool |73 |  
 |Logan |2017-02-28 |7.8 |Dans un futur proche, un certain  
 Logan, épuisé de fatigue, s'occupe d'un Professeur X souffrant, dans un li  
 eu gardé secret à la frontière Mexicaine. Mais les tentatives de Logan pou  
 r se retrancher du monde et rompre avec son passé vont s'épuiser lorsqu'un  
 e jeune mutante traquée par de sombres individus va se retrouver soudainem  
 ent face à lui.  
 |19250 |75.37 |Logan |en |Action, Dr  
 ame, Science-Fiction |Brazil |2017|Filmes |1 |Logan  
 |288 |  
 |Bohemian Rhapsody |2018-10-24 |7.984|Le parcours de Queen et son leade  
 r Freddie Mercury, de la formation du groupe à son apparition au concert L  
 ive Aid en 1985.  
 |16849 |42.576 |Bohemian Rhapsody|en |Musique, D  
 rame |Slovenia |2018|Filmi |1 |Bohem  
 ian Rhapsody |901 |  
 |Last Christmas |2019-11-07 |7.2 |Kate est une jeune femme abonnée  
 aux mauvaises décisions. Sa dernière en date ? Celle d'avoir acceptée de t  
 ravailer comme lutin du Père Noël pour un grand magasin. Pourtant, elle y  
 fait la rencontre de Tom. Sa vie prend une nouvelle tournure. Pour Kate, c  
 ela semble même trop beau pour être vrai !  
 |2358 |139.793 |Last Christmas |en |Comédie, R  
 omance |Croatia |2019|Filmovi |1 |Last  
 Christmas |176 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur proche, face à une  
 Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellair  
 e pour franchir un trou de ver permettant de parcourir des distances jusqu  
 e-là infranchissables. Leur but : trouver un nouveau foyer pour l'humanité  
 .  
 |35760 |238.105 |Interstellar |en |Aventure,  
 Drame, Science-Fiction |France |2014|Films |1 |Inter  
 stellar |21 |  
 |1917 |2019-12-25 |7.991|Pris dans la tourmente de la Prem  
 ière Guerre Mondiale, Schofield et Blake, deux jeunes soldats britanniques  
 , se voient assigner une mission à proprement parler impossible. Porteurs  
 d'un message qui pourrait empêcher une attaque dévastatrice et la mort de  
 centaines de soldats, dont le frère de Blake, ils se lancent dans une véri

table course contre la montre, derrière les lignes ennemies.

12378	35.96	1917	en	Guerre, Hi
stoire, Thriller, Drame	Netherlands	2020	Films & series	1  1917
1147				

|Joker |2019-10-01 |8.1 |Dans les années 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule dans la folie et devient le Joker.

25781	201.229	Joker	en	Crime, Thr
iller, Drame	Poland	2019	Filmy	1  Joker
56				

|Avengers : Endgame|2019-04-24 |8.2 |Après leur défaite face au Titan Thanos qui dans le film précédent s'est approprié toutes les pierres du Gauntlet de l'infini, les Avengers et les Gardiens de la Galaxie ayant survécu à son claquement de doigts qui a pulvérisé « la moitié de toute forme de vie dans l'Univers », Captain America, Thor, Bruce Banner, Natasha Romanoff, War Machine, Tony Stark, Nebula et Rocket, vont essayer de trouver une solution pour ramener leurs coéquipiers disparus et vaincre Thanos en se faisant aider par Ronin alias Clint Barton, Captain Marvel et Ant-Man.

25662	124.211	Avengers: Endgame	en	Aventure,
Science-Fiction, Action	France	2019	Les films	1  Aveng
ers : Endgame 109				

|Deadpool |2016-02-09 |7.624|Deadpool, est l'anti-héros le plus atypique de l'univers Marvel. À l'origine, il s'appelle Wade Wilson : un ancien militaire des Forces Spéciales devenu mercenaire. Après avoir subi une expérimentation hors norme qui va accélérer ses pouvoirs de guérison, il va devenir Deadpool. Armé de ses nouvelles capacités et d'un humour noir survolté, Deadpool va traquer l'homme qui a bien failli anéantir sa vie.

31055	190.413	Deadpool	en	Action, Av
enture, Comédie	Nigeria	2016	MOVIES / TV SERIES	1  Deadp
ool  78				

|Iron Man |2008-04-30 |7.647|Tony Stark, inventeur de génie, vendeur d'armes et playboy milliardaire, est kidnappé. Forcé par ses ravisseurs de fabriquer une arme redoutable, il construit en secret une armure high-tech révolutionnaire qu'il utilise pour s'échapper. Comprenant la puissance de cette armure, il décide de l'améliorer et de l'utiliser pour faire régner la justice et protéger les innocents.

26377	92.993	Iron Man	en	Action, Sc
ience-Fiction, Aventure	Global	2008	Movie Trailers	1  Iron
Man  123				

|Interstellar |2014-11-05 |8.4 |Dans un futur proche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellaire pour franchir un trou de ver permettant de parcourir des distances jusque-là infranchissables. Leur but : trouver un nouveau foyer pour l'humanité.

35760	238.105	Interstellar	en	Aventure,
Drame, Science-Fiction	United States	2014	Movies	1  Inter
stellar  15				

|Captain Marvel |2019-03-06 |6.8 |Captain Marvel raconte l'histoire de Carol Danvers qui va devenir l'une des super-héroïnes les plus puissantes de l'univers lorsque la Terre se révèle l'enjeu d'une guerre galactique entre deux races extraterrestres.

15685	58.773	Captain Marvel	en	Action, Av
enture, Science-Fiction	Myanmar (Burma)	2019	Movies & Series	1  Capta



Year	Genre	Avg_Popularity
2008	en	1797.5
2008	ja	96.0
2008	Science-Fiction	36.11454545454544
2008	Animation	35.304
2008	Science-Fiction	35.2205
2008	Action	33.184099999999999
2008	Action	33.053181818181805
2008	Fantastique	32.91975
2008	Aventure	32.34472413793104
2008	Thriller	31.952621621621624
2008	Guerre	31.072999999999997
2008	Crime	30.517444444444443
2008	Animation	30.167888888888886
2008	Romance	29.69959375
2008	Familial	29.29968965517241
2008	Mystère	29.023374999999998
2008	Drame	28.817928571428574
2008	Aventure	28.365466666666666
2008	Histoire	26.889857142857142
2008	Fantastique	26.451041666666667

only showing top 20 rows

## 7. Les films avec des résumés contenant des mots-clés spécifiques

```
In [22]: # Filtrer les films avec des résumés contenant des mots-clés
keyword_filtered = spark_dixk_filtered.filter(
    col("Résumé").rlike("(?i)aventure|magie|action")
)

keyword_filtered.show(truncate=False)
```

[illegible]



[illegible]

une lutte visant à maintenir la paix. Harold et Krokmu vont se battre pour défendre leurs valeurs et préserver le destin des hommes et des dragons.

9572.0	134.223	How to Train Your Dragon 2	e
n		Fantastique, Action, Aventure, Animation, Comédie, Familial NULL	NULL NULL

|Harry Potter et les Reliques de la mort – 1ère partie|2010-11-17 |7.74

7|Le pouvoir de Voldemort s'étend. Celui-ci contrôle maintenant le Ministère de la Magie et Poudlard. Harry, Ron et Hermione décident de terminer le travail commencé par Dumbledore, et de retrouver les derniers Horcruxes pour vaincre le Seigneur des Ténèbres. Mais il reste bien peu d'espoir aux trois sorciers, qui doivent réussir à tout prix.

19042.0	149.827	Harry Potter and the Deathly Hallows: Part 1 e
n		Aventure, Fantastique
NULL	NULL NULL	NULL NULL

|La Reine des neiges |2013-11-20 |7.24

5|Anna, une jeune fille aussi audacieuse qu'optimiste, se lance dans un incroyable voyage en compagnie de Kristoff, un montagnard expérimenté, et de son fidèle renne, Sven à la recherche de sa sœur, Elsa, la Reine des Neiges qui a plongé le royaume d'Arendelle dans un hiver éternel... En chemin, ils vont rencontrer de mystérieux trolls et un drôle de bonhomme de neige nommé Olaf, braver les conditions extrêmes des sommets escarpés et glacés, et affronter la magie qui les guette à chaque pas.

16502.0	134.957	Frozen	e
n		Animation, Familial, Aventure, Fantastique	
NULL	NULL NULL	NULL NULL	

|Pirates des Caraïbes : La Fontaine de jouvence |2011-05-15 |6.55

4|Dans cette histoire pleine d'action, où vérité, trahison, jeunesse éternelle et mort forment un cocktail explosif, le capitaine Jack Sparrow retrouve une femme qu'il a connu autrefois. Leurs liens sont-ils faits d'amour ou, cette femme n'est-elle qu'une aventurière sans scrupules qui cherche à l'utiliser pour découvrir la légendaire Fontaine de Jouvence ? Lorsqu'elle l'oblige à embarquer à bord du Queen Anne's Revenge, le bateau du terrible pirate Barbe-Noire, Jack ne sait plus ce qu'il doit craindre le plus : Le redoutable maître du bateau ou cette femme surgit de son passé...

13976.0	113.051	Pirates of the Caribbean: On Stranger Tides  e
n		Aventure, Action, Fantastique
NULL	NULL NULL	NULL NULL

|Inception |2010-07-15 |8.36

9|Dom Cobb est un voleur expérimenté, le meilleur dans l'art dangereux de l'extraction, voler les secrets les plus intimes enfouis au plus profond du subconscient durant une phase de rêve, lorsque l'esprit est le plus vulnérable. Les capacités de Cobb ont fait des envieux dans le monde tourmenté de l'espionnage industriel alors qu'il devient fugitif en perdant tout ce qu'il a un jour aimé. Une chance de se racheter lui est alors offerte. Une ultime mission grâce à laquelle il pourrait retrouver sa vie passée mais uniquement s'il parvient à accomplir l'impossible inception.

36635.0	122.2	Inception	e
n		Action, Science-Fiction, Aventure	
NULL	NULL NULL	NULL NULL	

|WALL•E |2008-06-22 |8.1

|Faites la connaissance de WALL•E (prononcez « Walli ») : WALL•E est le dernier être sur Terre et s'avère être un... petit robot ! 700 ans plus tôt, l'humanité a déserté notre planète laissant à cette incroyable petite machi

ne le soin de nettoyer la Terre. Mais au bout de ces longues années, WALL•E a développé un petit défaut technique : une forte personnalité. Extrêmement curieux, très indiscret, il est surtout un peu trop seul... Cependant, sa vie s'apprête à être bouleversée avec l'arrivée d'une petite « robote », bien carénée et prénommée EVE. Tombant instantanément et éperdument amoureux d'elle, WALL•E va tout mettre en œuvre pour la séduire. Et lorsqu'EVE est rappelée dans l'espace pour y terminer sa mission, WALL•E n'hésite pas un seul instant : il se lance à sa poursuite... Hors de question pour lui de laisser passer le seul amour de sa vie... Pour être à ses côtés, il est prêt à aller au bout de l'univers et vivre la plus fantastique des aventures !

18770.0 | 97.924 | WALL•E | en  
| Animation, Familial, Science-Fiction | NULL | NULL |  
NULL | NULL | NULL |

| Rio | 2011-04-03 | 6.7  
| Blu, un perroquet bleu d'une espèce très rare, quitte sa petite ville sous la neige et le confort de sa cage pour s'aventurer au cœur des merveilles exotiques de Rio de Janeiro. Sachant qu'il n'a jamais appris à voler, l'aventure grandiose qui l'attend au Brésil va lui faire perdre quelques plumes ! Heureusement, ses nouveaux amis hauts en couleurs sont prêts à tout réveiller le héros qui est en lui, et lui faire découvrir tout le sens de l'expression « prendre son envol ».

| 6642.0 | 172.696 | Rio | e  
n | Animation, Aventure, Comédie, Familial

| Mexico | 2011 | Movies | 2 | Rio |  
| Dragons : Retrouvailles | 2019-12-03 | 8.1  
| Quand il devient évident que la nouvelle génération de Vikings ne se souvient plus du lien entre dragon et humain, Hiccup définit un plan pour célébrer les dragons avec un grand spectacle de vacances de Snoggletog et obtient une visite surprise réconfortante de Toothless pour une autre aventure.

| 950.0 | 109.737 | How to Train Your Dragon: Homecoming | e  
n | Animation, Fantastique, Aventure, Action, Familial

| NULL | NULL | NULL | NULL | NULL |  
| Raiponce | 2010-11-24 | 7.6  
| Quand le bandit le plus recherché du royaume, Flynn Rider, se cache dans une mystérieuse tour, il est pris en otage par Raiponce, une jolie adolescente avec des cheveux de 20 mètres. La curieuse ravisseuse de Flynn, qui cherche à s'échapper de la tour où elle est enfermée depuis des années, conclut un accord avec le voleur et le duo improbable se lance dans une aventure pleine d'action, en compagnie du cheval Maximus, d'un camélon surprotecteur du nom de Pascal, et d'une bande de voyous.

| 11465.0 | 98.34 | Tangled | e  
n | Animation, Familial

| NULL | NULL | NULL | NULL | NULL |  
| La Princesse et la Grenouille | 2009-12-08 | 7.2  
| Cette version moderne d'un conte classique par Walt Disney Animation Studios a pour héroïne la jolie Tiana, qui rêve d'avoir son propre restaurant. Alors qu'elle rencontre un prince-grenouille désespéré de reprendre forme humaine, un baiser fatidique les entraîne tous deux dans une folle aventure au cours de laquelle le vilain magicien vaudou Dr Facilier les poursuit à travers les mystiques bayous de Louisiane.

| 5593.0 | 80.202 | The Princess and the Frog | e  
n | Animation, Romance, Fantastique, Familial

|NULL |NULL|NULL |NULL|NULL |  
 |Spider-Man : New Generation |2018-12-06 |8.39  
 9|Spider-Man : New Generation suit les aventures de Miles Morales, un adolescent afro-américain et portoricain qui vit à Brooklyn et s'efforce de s'intégrer dans son nouveau collège à Manhattan. Mais la vie de Miles se complique quand il se fait mordre par une araignée radioactive et se découvre des super-pouvoirs : il est désormais capable d'empoisonner ses adversaires, de se camoufler, de coller littéralement aux murs et aux plafonds ; son ouïe est démultipliée... Dans le même temps, le plus redoutable cerveau criminel de la ville, le Caïd, a mis au point un accélérateur de particules nucléaires capable d'ouvrir un portail sur d'autres univers. Son invention va provoquer l'arrivée de plusieurs autres versions de Spider-Man dans le monde de Miles, dont un Peter Parker plus âgé, Spider-Gwen, Spider-Man Noir, Spider-Cochon et Peni Parker, venue d'un dessin animé japonais.  
 |15655.0 |102.64 |Spider-Man: Into the Spider-Verse |e  
 n |Animation, Action, Aventure, Science-Fiction  
 |NULL |NULL|NULL |NULL|NULL |  
 |Là-haut |2009-05-28 |7.95  
 8|À 78 ans, Carl n'envisage plus qu'une chose : finir tranquillement sa vie dans la petite maison où il a passé toute une existence paisible auprès de son épouse Ellie. Veuf désormais, il reste, en effet, attaché à son passé en refusant les propositions mirobolantes d'un promoteur pour racheter son bien. Quand des circonstances exceptionnelles le contraignent d'intégrer une maison de retraite, il prend une ultime fois les choses en mains et, transformant sa demeure en dirigeable, s'envole pour l'aventure, vers la partie du monde où sa femme et lui s'étaient promis un jour de s'installer. Il ne se doute pas alors qu'il a emporté avec lui un jeune passager clandestin, Russell et que le voyage sera loin d'être de tout repos...  
 |20206.0 |86.637 |Up |e  
 n |Animation, Comédie, Familial, Aventure  
 |NULL |NULL|NULL |NULL|NULL |  
 |La liste de nos rêves |2018-12-05 |7.4  
 |Calvin est un jeune bagagiste qui a abandonné ses études et qui est légèrement hypocondriaque. Sa vie va changer lorsqu'il va faire la connaissance de Skye. En effet, cette jeune femme, qui a réalisé une liste de choses importantes à faire avant de mourir, va réussir à l'embarquer dans une aventure et lui faire oublier ses plus grandes peurs. Grâce à elle, il va apprendre ce que c'est de réellement vivre.  
 |476.0 |79.63 |Then Came You |e  
 n |Comédie, Drame  
 |NULL |NULL|NULL |NULL|NULL |  
 |Toy Story 3 |2010-06-16 |7.8  
 |Woody, Buzz l'Éclair et le reste des jouets reviennent dans une toute nouvelle aventure avec quelques nouveaux visages dont Ken et Barbie, le héros La Brosse et l'ours Lotso. Andy se prépare à partir pour l'université et Woody, Buzz, Jessie et les autres jouets se demandent ce qu'ils vont de venir. Quand ils arrivent à Sunnyside, c'est le début d'une nouvelle aventure.  
 |14647.0 |75.883 |Toy Story 3 |e  
 n |Animation, Familial, Comédie  
 |NULL |NULL|NULL |NULL|NULL |  
 |Alice au pays des merveilles |2010-03-03 |6.63  
 9|Alice, désormais âgée de 19 ans, retourne dans le monde fantastique qu'e

lle a découvert quand elle était enfant. Elle y retrouve ses amis le Lapin Blanc, Bonnet Blanc et Blanc Bonnet, le Loir, la Chenille, le Chat du Cheshire et, bien entendu, le Chapelier Fou. Alice s'embarque alors dans une aventure extraordinaire où elle accomplira son destin : mettre fin au règne de terreur de la Reine Rouge.

13929.0	83.225	Alice in Wonderland	e
n		Familial, Fantastique, Aventure	

NULL	NULL NULL	NULL NULL	
------	-----------	-----------	--

Ferdinand		2017-12-09	7.21
-----------	--	------------	------

|Ferdinand est un taureau au grand cœur. Victime de son imposante apparence, il se retrouve malencontreusement capturé et arraché à son village d'origine. Bien déterminé à retrouver sa famille et ses racines, il se lance alors dans une incroyable aventure à travers l'Espagne, accompagné de la plus déjantée des équipes !

2836.0	113.396	Ferdinand	e
n		Animation, Familial, Aventure, Comédie	

NULL	NULL NULL	NULL NULL	
------	-----------	-----------	--

L'Âge de glace : Les Lois de l'Univers		2016-06-23	6.1
--	--	------------	-----

|La quête permanente de Scratt pour attraper son insaisissable noisette le catapulte dans l'espace où il déclenche accidentellement une série d'événements cosmiques qui vont transformer et menacer le monde de l'Âge de Glace. Pour sauver leur peau, Sid, Manny, Diego et le reste de la bande vont devoir quitter leur foyer et embarquer dans une nouvelle aventure pleine de comédie, parcourant de nouvelles terres exotiques et rencontrant une galerie de personnages hauts en couleur.

4531.0	67.368	Ice Age: Collision Course	e
n		Aventure, Animation, Familial, Comédie, Science-Fiction	

NULL	NULL NULL	NULL NULL	
------	-----------	-----------	--

Madagascar 3 : Bons baisers d'Europe		2012-06-06	6.61
--------------------------------------	--	------------	------

2|Alex le lion, Marty le zèbre, Gloria l'hippopotame et Melman la girafe cherchent toujours désespérément à rentrer chez eux ! Évidemment, King Julien, Maurice et les pingouins les suivent dans leurs péripéties... Leurs aventures les mènent en Europe où ils trouvent une couverture idéale : un cirque ambulancier, qu'ils vont devoir ré-inventer – façon Madagascar...

6109.0	112.643	Madagascar 3: Europe's Most Wanted	e
n		Animation, Familial, Comédie, Aventure	

NULL	NULL NULL	NULL NULL	
------	-----------	-----------	--

+	+	+	+
---	---	---	---

+			
---	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

--	--	--	--

```
-----+-----+-----+-----+-----+
only showing top 20 rows
```

## 8. Les films dont la popularité dépasse la moyenne du DataFrame

```
In [23]: from pyspark.sql.functions import mean

# Calculer la moyenne de popularité
average_popularity = spark_dixk_filtered.agg(mean("Popularité").alias("Av

# Filtrer les films dont la popularité dépasse la moyenne
above_average_movies = spark_dixk_filtered.filter(col("Popularité") > ave

above_average_movies.show(truncate=False)
```

```
+-----+-----+-----+-----+-----+
|titre|Date de sortie|Note|Résumé|
|Nombre de votes|Popularité|original title|Original Lan|
|guage|Genres|Country|Year|
|Category|Rank|Query|
+-----+-----+-----+-----+-----+
|Le Grinch|2018-11-08|6.864|Terré dans sa cave|
|remplie d'inventions et d'engins pour ses besoins de tous les jours, Le Gr|
|inch ne voit ses voisins de Chouville que lorsqu'il manque de nourriture.|
|Chaque année à Noël, ils viennent perturber sa tranquillité solitaire avec|
|des célébrations toujours plus grandioses, brillantes et bruyantes. Quand|
|les Chous déclarent qu'ils vont célébrer Noël trois fois plus fort cette a|
|nnée, le Grinch réalise qu'il n'a plus qu'une solution pour retrouver la p|
|aix et la tranquillité : il doit voler Noël.|3944.0|477.459|Th| |
|e Grinch|en|Animation, Familial, Com|
|édie, Fantastique|NULL|NULL|NULL|NULL|
|
```

|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |United Kingdom |2018|Movie  
s |5 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |Singapore |2018|Top T  
rending Movies|5 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |Romania |2018|Filme  
|3 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |Nigeria |2018|Movie  
s |4 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |Malaysia |2018|Filem  
|3 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |Hungary |2018|Filme  
k |1 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !  
|16065.0 |301.84 |Venom |en  
|Science-Fiction, Action |France |2018|Les f  
ilms |5 |Venom |  
|Venom |2018-09-28 |6.8 |Des symbiotes déb  
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D  
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave  
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !

16065.0	301.84	Venom	en
Science-Fiction, Action	Canada	2018 Movie	
s	4	Venom	
Venom	2018-09-28	6.8	Des symbiotes déba
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D			
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave			
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !			
16065.0	301.84	Venom	en
Science-Fiction, Action	Brazil	2018 Filme	
s	5	Venom	
Venom	2018-09-28	6.8	Des symbiotes déba
rquent sur la Terre, parmi eux, Venom, qui va s'allier avec Eddie Brock. D			
e son côté, un puissant scientifique tente de faire évoluer l'humanité ave			
c les symbiotes, le duo d'anti-héros va devoir tout faire pour l'arrêter !			
16065.0	301.84	Venom	en
Science-Fiction, Action	Global	2018 Movie	
s	3	Venom	
일본 엄마	2017-02-09	6.2	NULL
28.0	364.704	일본 엄마	ko
Romance, Drame	NULL	NULL NULL	
NULL NULL			
Des jours plus belles que la nuit 2010-02-19	6.523	Wim et Floor, un j	
eune couple en chair et en os, passent un après-midi dans les pièces ensol			
eillées d'une maison en ruine en Belgique. Ce documentaire artistique érot			
ique est monté presque en temps réel. Dans la lenteur, nous obtenons la co			
nstruction, la douceur et la sexualité. Oubliez les montages rapides qui c			
laquent des doigts. C'est vers la lenteur qu'il faut aller.			
22.0	299.557	Des jours plus belles que la nuit nl	
Romance	NULL	NULL NULL	
NULL NULL			
Avengers : Infinity War	2018-04-25	8.2	Les Avengers et le
urs alliés devront être prêts à tout sacrifier pour neutraliser le redouta			
ble Thanos avant que son attaque éclair ne conduise à la destruction compl			
ète de l'univers.			
29791.0	342.224	Avengers: Infinity War	en
Aventure, Action, Science-Fiction	NULL	NULL NULL	
NULL NULL			
Equalizer	2014-09-24	7.281	McCall, un homme q
ui pense avoir rangé son passé mystérieux derrière lui, se consacre à sa n			
ouvelle vie tranquille. Au moment où il rencontre Teri, une jeune fille so			
us le contrôle de gangsters russes violents, il décide d'agir. McCall sort			
ainsi de sa retraite et voit son désir de justice réveillé.			
9029.0	192.911	The Equalizer	en
Thriller, Action, Crime	NULL	NULL NULL	
NULL NULL			
Interstellar	2014-11-05	8.4	Dans un futur proc
he, face à une Terre exsangue, un groupe d'explorateurs utilise un vaissea			
u interstellaire pour franchir un trou de ver permettant de parcourir des			
distances jusque-là infranchissables. Leur but : trouver un nouveau foyer			
pour l'humanité.			
35760.0	238.105	Interstellar	en
Aventure, Drame, Science-Fiction	United States	2014 Movie	
s	2	Interstellar	



Interstellar	2014-11-05	8.4	Dans un futur proc
he, face à une Terre exsangue, un groupe d'explorateurs utilise un vaissea			
u interstellaire pour franchir un trou de ver permettant de parcourir des			
distances jusque-là infranchissables. Leur but : trouver un nouveau foyer			
pour l'humanité.			
35760.0	238.105	Interstellar	en
Aventure, Drame, Science-Fiction		United Arab Emirates 2014 Movie	
s	2	Interstellar	
Interstellar	2014-11-05	8.4	Dans un futur proc
he, face à une Terre exsangue, un groupe d'explorateurs utilise un vaissea			
u interstellaire pour franchir un trou de ver permettant de parcourir des			
distances jusque-là infranchissables. Leur but : trouver un nouveau foyer			
pour l'humanité.			
35760.0	238.105	Interstellar	en
Aventure, Drame, Science-Fiction		Sweden	2014 Film
1	Interstellar		
Interstellar	2014-11-05	8.4	Dans un futur proc
he, face à une Terre exsangue, un groupe d'explorateurs utilise un vaissea			
u interstellaire pour franchir un trou de ver permettant de parcourir des			
distances jusque-là infranchissables. Leur but : trouver un nouveau foyer			
pour l'humanité.			
35760.0	238.105	Interstellar	en
Aventure, Drame, Science-Fiction		Singapore	2014 Movie
s	1	Interstellar	
Interstellar	2014-11-05	8.4	Dans un futur proc
he, face à une Terre exsangue, un groupe d'explorateurs utilise un vaissea			
u interstellaire pour franchir un trou de ver permettant de parcourir des			
distances jusque-là infranchissables. Leur but : trouver un nouveau foyer			
pour l'humanité.			
35760.0	238.105	Interstellar	en
Aventure, Drame, Science-Fiction		Malaysia	2014 Movie
s (Film)	3	Interstellar	

only showing top 20 rows

## 9. Obtenir des statistiques descriptives des notes

```
In [24]: # Obtenir des statistiques descriptives des notes, Popularité et Nombres
spark_dixk_filtered.describe("Note","Popularité","Nombre de votes").show()
```

summary	Note	Popularité	Nombre de votes
count	4077	3957	3928
mean	6.452360068677949	49.811545867283655	3196.6970468431773
stddev	1.0096664068131147	358.86337768748604	5580.345161145155
min	0.0	9.628	0.0
max	9.8	20957.0	36635.0

## 10. Film avec la note la plus haute, plus de votes et populaires

```
In [25]: from pyspark.sql.functions import col, max
from pyspark.sql.types import FloatType

# Convertir Popularité et Nombre de votes en FloatType
spark_dixk_filtered_10 = spark_dixk_filtered.withColumn("Popularité", col("Popularité").cast(FloatType))
                                              .withColumn("Nombre de votes", col("Nombre de votes").cast(FloatType))

# Calculer les valeurs maximales
max_values = spark_dixk_filtered_10.agg(
    max(col("Note")).alias("Max_Note"),
    max(col("Nombre de votes")).alias("Max_Votes"),
    max(col("Popularité")).alias("Max_Popularity")
).collect()[0]

max_note = max_values["Max_Note"]
max_votes = max_values["Max_Votes"]
max_popularity = max_values["Max_Popularity"]

print(f"Max Note: {max_note}, Max Votes: {max_votes}, Max Popularity: {max_popularity}")
```

Max Note: 9.8, Max Votes: 36635.0, Max Popularity: 20957.0

filtré que par popularité ça pue la merde pour faire des comparaisons, c'est trop biaisés et mal renseignés et uniquement par 1 seul n'est par pertinent.

## Rank pour pour afficher les données

```
In [26]: from pyspark.sql.window import Window
from pyspark.sql.functions import rank, col
from pyspark.sql.functions import max
from pyspark.sql.types import FloatType

# Étape 1 : Convertir les colonnes en types numériques si nécessaire
spark_dixk_filtered = spark_dixk_filtered.withColumn("Popularité", col("Popularité").cast(FloatType))
                                              .withColumn("Nombre de votes", col("Nombre de votes").cast(FloatType))

# Étape 2 : Définir des fenêtres pour attribuer un rang
```

```

window = Window.orderBy(col("Note").desc())
window_votes = Window.orderBy(col("Nombre de votes").desc())
window_popularity = Window.orderBy(col("Popularité").desc())

# Étape 3 : Ajouter les rangs pour chaque critère
ranked_movies = spark_dixk_filtered \
    .withColumn("Rank_Note", rank().over(window)) \
    .withColumn("Rank_Votes", rank().over(window_votes)) \
    .withColumn("Rank_Popularity", rank().over(window_popularity))

# Étape 4 : Calculer un score global en combinant les rangs
ranked_movies = ranked_movies.withColumn(
    "Overall_Rank", col("Rank_Note") + col("Rank_Votes") + col("Rank_Popularity")
)

# Étape 5 : Trier par le score global (plus bas est meilleur)
ranked_movies_sorted = ranked_movies.orderBy(col("Overall_Rank"))

# Étape 6 : Trier par le score global et sélectionner le Top 10
ranked_movies_sorted = ranked_movies.orderBy(col("Overall_Rank")).limit(20)

# Étape 7 : Afficher les résultats
ranked_movies_sorted.show(truncate=False)

```

```

+-----+-----+-----+-----+
|titre|Date de sortie|Note|Résumé|
|Nombre de votes|Popularité|original title|Original Language|Genre|
s|Country|Year|Category|Rank|
k|Query|Rank_Note|Rank_Votes|Rank_Popularity|Overall_Rank|
+-----+-----+-----+-----+
|Interstellar|2014-11-05|8.4|Dans un futur pr
oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss

```

eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |Malaysia |2014|Movies (Filem) |3  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |United States |2014|Movies |2  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |United Arab Emirates|2014|Movies |2  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |Sweden |2014|Film |1  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |Singapore |2014|Movies |1  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |Ireland |2014|Movies |5  
 |Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.  
 |35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |France |2014|Films |2

|Interstellar|13 |2 |17 |32 |  
 |Interstellar |2014-11-05 |8.4 |Dans un futur pr  
 oche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaiss  
 eau interstellaire pour franchir un trou de ver permettant de parcourir de  
 s distances jusque-là infranchissables. Leur but : trouver un nouveau foye  
 r pour l'humanité.

|35760.0 |238.105 |Interstellar |en |Avent  
 ure, Drame, Science-Fiction |Canada |2014|Movies |2  
 |Interstellar|13 |2 |17 |32 |  
 |Avengers : Infinity War |2018-04-25 |8.2 |Les Avengers et  
 leurs alliés devront être prêts à tout sacrifier pour neutraliser le redou  
 table Thanos avant que son attaque éclair ne conduise à la destruction com  
 plète de l'univers.

|29791.0 |342.224 |Avengers: Infinity War|en |Avent  
 ure, Action, Science-Fiction|NULL |NULL|NULL |NUL  
 L|NULL |46 |30 |3 |79 |  
 |The Dark Knight : Le Chevalier noir|2008-07-16 |8.516|Batman aborde un  
 e phase décisive de sa guerre contre le crime à Gotham City. Avec l'aide d  
 u lieutenant de police Jim Gordon et du nouveau procureur Harvey Dent, il  
 entreprend de démanteler les dernières organisations criminelles qui infes  
 tent les rues de la ville. L'association s'avère efficace, mais le trio se  
 heurte bientôt à un nouveau génie du crime qui répand la terreur et le cha  
 os dans Gotham : le Joker. On ne sait pas d'où il vient ni qui il est. Ce  
 criminel possède une intelligence redoutable doublé d'un humour sordide et  
 n'hésite pas à s'attaquer à la pègre locale dans le seul but de semer le c  
 haos.|32971.0 |142.229 |The Dark Knight |en |  
 Drame, Action, Crime, Thriller |NULL |NULL|NULL  
 |NULL|NULL |7 |10 |97 |114 |  
 |Joker |2019-10-01 |8.1 |Dans les années  
 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule  
 dans la folie et devient le Joker.

|25781.0 |201.229 |Joker |en |Crime  
 , Thriller, Drame |Singapore |2019|Trending Movies|3  
 |Joker |79 |41 |27 |147 |  
 |Joker |2019-10-01 |8.1 |Dans les années  
 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule  
 dans la folie et devient le Joker.

|25781.0 |201.229 |Joker |en |Crime  
 , Thriller, Drame |Indonesia |2019|Film |1  
 |Joker |79 |41 |27 |147 |  
 |Joker |2019-10-01 |8.1 |Dans les années  
 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule  
 dans la folie et devient le Joker.

|25781.0 |201.229 |Joker |en |Crime  
 , Thriller, Drame |Poland |2019|Filmy |2  
 |Joker |79 |41 |27 |147 |  
 |Joker |2019-10-01 |8.1 |Dans les années  
 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule  
 dans la folie et devient le Joker.

|25781.0 |201.229 |Joker |en |Crime  
 , Thriller, Drame |Germany |2019|Filme |1  
 |Joker |79 |41 |27 |147 |  
 |Joker |2019-10-01 |8.1 |Dans les années

```
from pyspark.sql.window import Window
from pyspark.sql.functions import row_number, col

# Étape 1 : Convertir les colonnes en types numériques si nécessaire
spark_dixk_filtered = spark_dixk_filtered.withColumn("Popularité", col("P
```

```
.withColumn("Nombre de votes", c

# Étape 2 : Définir des fenêtres pour attribuer un rang
window_note = Window.orderBy(col("Note").desc())
window_votes = Window.orderBy(col("Nombre de votes").desc())
window_popularity = Window.orderBy(col("Popularité").desc())

# Ajouter les rangs pour chaque critère
ranked_movies = spark_dixk_filtered \
    .withColumn("Rank_Note", rank().over(window_note)) \
    .withColumn("Rank_Votes", rank().over(window_votes)) \
    .withColumn("Rank_Popularity", rank().over(window_popularity))

# Calculer un score global en combinant les rangs
ranked_movies = ranked_movies.withColumn(
    "Overall_Rank", col("Rank_Note") + col("Rank_Votes") + col("Rank_Popu
)

# Étape 3 : Limiter à une seule observation par titre
# Définir une fenêtre partitionnée par titre, triée par Overall_Rank
window_unique_title = Window.partitionBy("titre").orderBy(col("Overall_Ra

# Ajouter un numéro de ligne pour chaque partition
unique_movies = ranked_movies.withColumn("row_num", row_number().over(win

# Filtrer pour ne conserver que la première ligne pour chaque titre
unique_movies_filtered = unique_movies.filter(col("row_num") == 1)

# Étape 4 : Trier les résultats finaux par Overall_Rank et limiter au top
top_10_movies = unique_movies_filtered.orderBy(col("Overall_Rank")).limit

# Étape 5 : Afficher les résultats
top_10_movies.show(truncate=False)
```

[illegible]

original title	Country	Year	Category	Rank	Query	Rank_Note	Rank_Votes	Rank_Popularity	Overall_Rank	row_num
Interstellar	United States	2014	Movies	2	Interstellar	13	2	17	32	1
<p>Dans un futur proche, face à une Terre exsangue, un groupe d'explorateurs utilise un vaisseau interstellaire pour franchir un trou de ver permettant de parcourir des distances jusque-là infranchissables. Leur but : trouver un nouveau foyer pour l'humanité.</p>										
Avengers : Infinity War	United States	2018	Movies	1	Avengers: Infinity War	46	30	3	79	1
<p>Les Avengers et leurs alliés devront être prêts à tout sacrifier pour neutraliser le redoutable Thanos avant que son attaque éclair ne conduise à la destruction complète de l'univers.</p>										
The Dark Knight : Le Chevalier noir	United States	2008	Movies	6	The Dark Knight	7	10	97	114	1
<p>Batman aborde une phase décisive de sa guerre contre le crime à Gotham City. Avec l'aide du lieutenant de police Jim Gordon et du nouveau procureur Harvey Dent, il entreprend de démanteler les dernières organisations criminelles qui infestent les rues de la ville. L'association s'avère efficace, mais le trio se heurte bientôt à un nouveau génie du crime qui répand la terreur et le chaos dans Gotham : le Joker. On ne sait pas d'où il vient ni qui il est. Ce criminel possède une intelligence redoutable doublée d'un humour sordide et n'hésite pas à s'attaquer à la pègre locale dans le seul but de semer le chaos.</p>										



|Joker |2019-10-01 |8.1  
 |Dans les années 1980, à Gotham City, Arthur Fleck, un humoriste de stand-up raté, bascule dans la folie et devient le Joker.  
 |25781.0 |201.229 |Joker |e  
 n |Crime, Thriller, Drame |United Sta  
 tes |2019|Movies |3 |Joker |79 |41 |27  
 |147 |1 |  
 |Inception |2010-07-15 |8.36  
 9|Dom Cobb est un voleur expérimenté, le meilleur dans l'art dangereux de l'extraction, voler les secrets les plus intimes enfouis au plus profond du subconscient durant une phase de rêve, lorsque l'esprit est le plus vulnérable. Les capacités de Cobb ont fait des envieux dans le monde tourmenté de l'espionnage industriel alors qu'il devient fugitif en perdant tout ce qu'il a un jour aimé. Une chance de se racheter lui est alors offerte. Une ultime mission grâce à laquelle il pourrait retrouver sa vie passée mais uniquement s'il parvient à accomplir l'impossible inception.  
 |36635.0 |122.2 |Inception |e  
 n |Action, Science-Fiction, Aventure |NULL  
 |NULL|NULL |NULL|NULL |32 |1 |136  
 |169 |1 |  
 |Coco |2017-10-27 |8.2  
 |Depuis déjà plusieurs générations, la musique est bannie dans la famille de Miguel. Un vrai déchirement pour le jeune garçon dont le rêve ultime est de devenir un musicien aussi accompli que son idole, Ernesto de la Cruz. Bien décidé à prouver son talent, Miguel, par un étrange concours de circonstances, se retrouve propulsé dans un monde étonnant et coloré lié aux célébrations des ancêtres. Là, il se lie d'amitié avec Hector, un gentil garçon mais un peu filou sur les bords. Ensemble, ils vont accomplir un voyage extraordinaire qui leur révélera la véritable histoire qui se cache derrière celle de la famille de Miguel..  
 |19514.0 |324.72 |Coco |e  
 n |Familial, Animation, Musique, Aventure |NULL  
 |NULL|NULL |NULL|NULL |46 |134 |4  
 |184 |1 |  
 |Avengers : Endgame |2019-04-24 |8.2  
 |Après leur défaite face au Titan Thanos qui dans le film précédent s'est approprié toutes les pierres du Gant de l'infini, les Avengers et les Gardiens de la Galaxie ayant survécu à son claquement de doigts qui a pulvérisé « la moitié de toute forme de vie dans l'Univers », Captain America, Thor, Bruce Banner, Natasha Romanoff, War Machine, Tony Stark, Nébula et Rocket, vont essayer de trouver une solution pour ramener leurs coéquipiers disparus et vaincre Thanos en se faisant aider par Ronin alias Clint Barton, Captain Marvel et Ant-Man.  
 |25662.0 |124.211 |Avengers: Endgame |e  
 n |Aventure, Science-Fiction, Action |France  
 |2019|Les films|2 |Avengers : Endgame|46 |58 |131  
 |235 |1 |  
 |Django Unchained |2012-12-25 |8.2  
 |Dans le sud des États-Unis, deux ans avant la guerre de Sécession, le Dr King Schultz, un chasseur de primes allemand, fait l'acquisition de Django, un esclave qui peut l'aider à traquer les frères Brittle, les meurtriers qu'il recherche. Schultz promet à Django de lui rendre sa liberté lorsqu'il aura capturé les Brittle – morts ou vifs. Alors que les deux hommes pist

ent les dangereux criminels, Django n'oublie pas que son seul but est de retrouver Broomhilda, sa femme, dont il fut séparé à cause du commerce des esclaves... Lorsque Django et Schultz arrivent dans l'immense plantation du puissant Calvin Candie, ils éveillent les soupçons de Stephen, un esclave qui sert Candie et a toute sa confiance. Le moindre de leurs mouvements est désormais épié par une dangereuse organisation de plus en plus proche... Si Django et Schultz veulent espérer s'enfuir avec Broomhilda, ils vont devoir choisir entre l'indépendance et la solidarité, entre le sacrifice et la survie...|26237.0 |103.622 |Django Unchained

|en |Drame, Western |Switzerland  
and |2013|Filme |1 |Django Unchained |46 |34 |178  
|258 |1 |

|기생충 |2019-05-30 |8.506

|Toute la famille de Ki-taek est au chômage. Elle s'intéresse particulièrement au train de vie de la riche famille Park. Mais un incident se produit et les deux familles se retrouvent mêlées, sans le savoir, à une bien étrange histoire...

|18347.0 |110.555 |Parasite |k  
o |Comédie, Thriller, Drame |NULL

|NULL|NULL |NULL|NULL |8 |150 |152

|310 |1 |

|Harry Potter et les Reliques de la mort - 2ème partie|2011-07-12 |8.09

2|Le combat entre les puissances du bien et du mal de l'univers des sorciers se transforme en guerre sans merci. L'école Poudlard est désormais sous la coupe du professeur Rogue et des forces des ténèbres. Les enjeux sont devenus considérables et personne n'est plus en sécurité alors que se rapproche l'ultime épreuve de force avec Voldemort...

|20547.0 |143.398 |Harry Potter and the Deathly Hallows: Part 2|e  
n |Fantastique, Aventure |NULL

|NULL|NULL |NULL|NULL |114 |108 |96

|318 |1 |

|Le Loup de Wall Street |2013-12-25 |8.0

|Sa licence de courtier en poche, et les narines déjà pleines de cocaïne, Jordan Belfort est prêt à conquérir Wall Street. Ce jour d'octobre, un krach, le plus important depuis 1929, vient piétiner ses rêves de grandeur. C'est finalement à Long Island qu'il échoue et qu'il monte sa propre affaire de courtage. Son créneau : le hors-cote. Sa méthode : l'arnaque. Son équipe : des vendeurs ou des petits trafiquants.

|23968.0 |119.995 |The Wolf of Wall Street |e  
n |Crime, Drame, Comédie |NULL

|NULL|NULL |NULL|NULL |133 |60 |138

|331 |1 |

|Vice-versa |2015-06-17 |7.9

|Riley est déracinée du Mid-Ouest quand son père est muté à San Francisco. Comme tous, Riley est guidée par ses Émotions : Joie, Peur, Colère, Dégoût et Tristesse. Les Émotions vivent au Quartier Cérébral, où elles la conseillent au quotidien. Alors que Riley et ses Émotions se bagarrent, c'est la crise au Quartier Cérébral. Joie, l'Émotion principale de Riley, essaie de rester positive alors que les autres Émotions essaient de s'adapter à la nouvelle vie de Riley.

|21988.0 |126.962 |Inside Out |e  
n |Animation, Familial, Aventure, Drame, Comédie |NULL

|NULL|NULL |NULL|NULL |180 |88 |129

|397 |1 |  
 |Deadpool |2016-02-09 |7.62  
 4|Deadpool, est l'anti-héros le plus atypique de l'univers Marvel. À l'origine, il s'appelle Wade Wilson : un ancien militaire des Forces Spéciales devenu mercenaire. Après avoir subi une expérimentation hors norme qui va accélérer ses pouvoirs de guérison, il va devenir Deadpool. Armé de ses nouvelles capacités et d'un humour noir survolté, Deadpool va traquer l'homme qui a bien failli anéantir sa vie.  
 |31055.0 |190.413 |Deadpool |e  
 n |Action, Aventure, Comédie |United Kin  
 gdom|2016|Movies |1 |Deadpool |351 |12 |46  
 |409 |1 |  
 |WALL•E |2008-06-22 |8.1  
 |Faites la connaissance de WALL•E (prononcez « Walli ») : WALL•E est le dernier être sur Terre et s'avère être un... petit robot ! 700 ans plus tôt, l'humanité a déserté notre planète laissant à cette incroyable petite machine le soin de nettoyer la Terre. Mais au bout de ces longues années, WALL•E a développé un petit défaut technique : une forte personnalité. Extrêmement curieux, très indiscret, il est surtout un peu trop seul... Cependant, sa vie s'apprête à être bouleversée avec l'arrivée d'une petite « robote », bien carénée et prénommée EVE. Tombant instantanément et éperdument amoureux d'elle, WALL•E va tout mettre en œuvre pour la séduire. Et lorsqu'EVE est rappelée dans l'espace pour y terminer sa mission, WALL•E n'hésite pas un seul instant : il se lance à sa poursuite... Hors de question pour lui de laisser passer le seul amour de sa vie... Pour être à ses côtés, il est prêt à aller au bout de l'univers et vivre la plus fantastique des aventures !  
 |18770.0 |97.924 |WALL•E |e  
 n |Animation, Familial, Science-Fiction |NULL  
 |NULL|NULL |NULL|NULL |79 |142 |206  
 |427 |1 |  
 |Spider-Man : New Generation |2018-12-06 |8.39  
 9|Spider-Man : New Generation suit les aventures de Miles Morales, un adolescent afro-américain et portoricain qui vit à Brooklyn et s'efforce de s'intégrer dans son nouveau collège à Manhattan. Mais la vie de Miles se complique quand il se fait mordre par une araignée radioactive et se découvre des super-pouvoirs : il est désormais capable d'empoisonner ses adversaires, de se camoufler, de coller littéralement aux murs et aux plafonds ; son ouïe est démultipliée... Dans le même temps, le plus redoutable cerveau criminel de la ville, le Caïd, a mis au point un accélérateur de particules nucléaires capable d'ouvrir un portail sur d'autres univers. Son invention va provoquer l'arrivée de plusieurs autres versions de Spider-Man dans le monde de Miles, dont un Peter Parker plus âgé, Spider-Gwen, Spider-Man Noir, Spider-Cochon et Peni Parker, venue d'un dessin animé japonais.  
 |15655.0 |102.64 |Spider-Man: Into the Spider-Verse |e  
 n |Animation, Action, Aventure, Science-Fiction |NULL  
 |NULL|NULL |NULL|NULL |27 |224 |186  
 |437 |1 |  
 |Gone Girl |2014-10-01 |7.89  
 2|Amy et Nick forment en apparence un couple modèle. Victimes de la crise, ils quittent Manhattan pour retourner s'installer dans la ville du Missouri où Nick a grandi. Mais le jour de leur 5ème anniversaire de mariage, Amy disparaît et Nick retrouve leur maison saccagée. Lors de l'enquête tout semble accuser Nick. Celui-ci décide, de son côté, de tout faire pour savoir

ce qui est arrivé à Amy et découvre qu'elle lui dissimulait beaucoup de choses.

18586.0	146.067	Gone Girl				
n		Mystère, Thriller, Drame			United States	
2014	Movies	4	Gone Girl	211	145	82
438	1					

Avengers		2012-04-25	7.72
----------	--	------------	------

3|Lorsque la sécurité et l'équilibre de la planète sont menacés par un ennemi d'un genre nouveau, Nick Fury, le directeur du SHIELD, l'agence internationale du maintien de la paix, réunit une équipe pour empêcher le monde de basculer dans le chaos. Partout sur Terre, le recrutement des nouveaux héros dont le monde a besoin commence...

30816.0	119.893	The Avengers				
n		Science-Fiction, Action, Aventure			Poland	
2012	Movies	4	Avengers	286	29	139
454	1					

君の名は。		2016-08-26	8.485
-------	--	------------	-------

Mitsuha est une lycéenne, la fille du maire d'une petite ville nichée entre les montagnes. Vivant avec sa petite sœur et sa grand-mère, c'est une demoiselle franche qui n'hésite pas à dire qu'elle n'a pas envie de participer aux rituels shinto, ou d'aider son père dans ses campagnes électorales. En fait, elle rêve de pouvoir quitter cette ville où elle s'ennuie, pour partir tenter sa chance à la capitale. Taki est un lycéen, un tokyoïte qui travaille à mi-temps dans un restaurant italien, tout en aspirant à des études d'architecture ou dans les Beaux-Arts. Chaque nuit, il fait un rêve étrange, où il devient... une lycéenne campagnarde, vivant dans une petite ville entre les montagnes. Quel secret se cache derrière ces rêves mutuels, qui unissent ces deux êtres qui ne se sont jamais rencontrés ?

11406.0	118.19	Your Name.				
a		Animation, Romance, Drame			NULL	
NULL	NULL	NULL	NULL	10	343	142
495	1					

Harry Potter et les Reliques de la mort - 1ère partie	2010-11-17	7.74
---	------------	------

7|Le pouvoir de Voldemort s'étend. Celui-ci contrôle maintenant le Ministère de la Magie et Poudlard. Harry, Ron et Hermione décident de terminer le travail commencé par Dumbledore, et de retrouver les derniers Horcruxes pour vaincre le Seigneur des Ténèbres. Mais il reste bien peu d'espoir aux trois sorciers, qui doivent réussir à tout prix.

19042.0	149.827	Harry Potter and the Deathly Hallows: Part 1				
n		Aventure, Fantastique			NULL	
NULL	NULL	NULL	NULL	279	140	79
498	1					

Avatar		2009-12-15	7.58
--------	--	------------	------

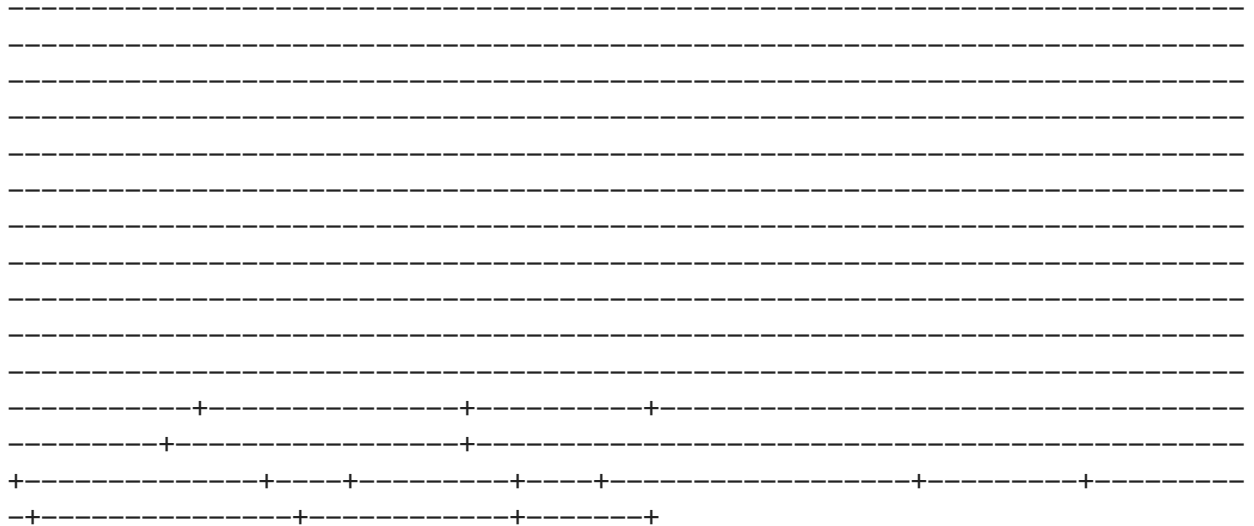
3|Un marine paraplégique, envoyé sur la lune Pandora pour une mission unique, est tiraillé entre suivre ses ordres et protéger le monde qu'il considère dorénavant comme le sien.

31545.0	187.204	Avatar				
n		Action, Aventure, Fantastique, Science-Fiction			NULL	
NULL	NULL	NULL	NULL	426	11	63
500	1					

```

+-----+-----+-----+
+-----+-----+-----+
+-----+-----+-----+

```



## 28/01/2025 : création de graphiques

### Code en Pandas et no pyspark

In [28]: `!pip install networkx matplotlib`

```
Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packages (3.4.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (4.55.7)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)
```

### graphes en 1 code

```

In [16]: import matplotlib.pyplot as plt
import networkx as nx
from collections import Counter

pdf = spark_dixk_filtered.toPandas()

# Construire les relations uniquement entre les genres
genre_edges = []

for _, row in pdf.iterrows():
    genres = row["Genres"].split(", ") if isinstance(row["Genres"], str)
    for i, genre1 in enumerate(genres):
        for genre2 in genres[i + 1:]:
            genre_edges.append((genre1, genre2))

# Compter les relations pour pondérer les arêtes
edge_weights = Counter(genre_edges)

# Créer un graphe à partir des relations entre les genres
G_genres = nx.Graph()
for edge, weight in edge_weights.items():
    G_genres.add_edge(edge[0], edge[1], weight=weight)

# Disposition des nœuds pour éviter les chevauchements
pos = nx.spring_layout(G_genres, k=0.5, seed=42)

# Taille des nœuds proportionnelle à leur degré
node_sizes = [G_genres.degree(n) * 150 for n in G_genres.nodes()]

# Couleur des nœuds basée sur le degré (nombre de connexions)
node_degrees = [G_genres.degree(n) for n in G_genres.nodes()]
cmap = plt.cm.Blues # Palette de couleurs

# Épaisseur des arêtes proportionnelle au poids (nombre de co-occurrences)
edge_widths = [G_genres[u][v]['weight'] * 0.2 for u, v in G_genres.edges()]

# Taille de la figure pour une meilleure lisibilité
plt.figure(figsize=(20, 15))

# Dessiner le graphe
nodes = nx.draw_networkx_nodes(
    G_genres,
    pos,
    node_size=node_sizes,
    node_color=node_degrees,
    cmap=cmap,
    alpha=0.9
)
nx.draw_networkx_edges(
    G_genres,
    pos,
    width=edge_widths,

```

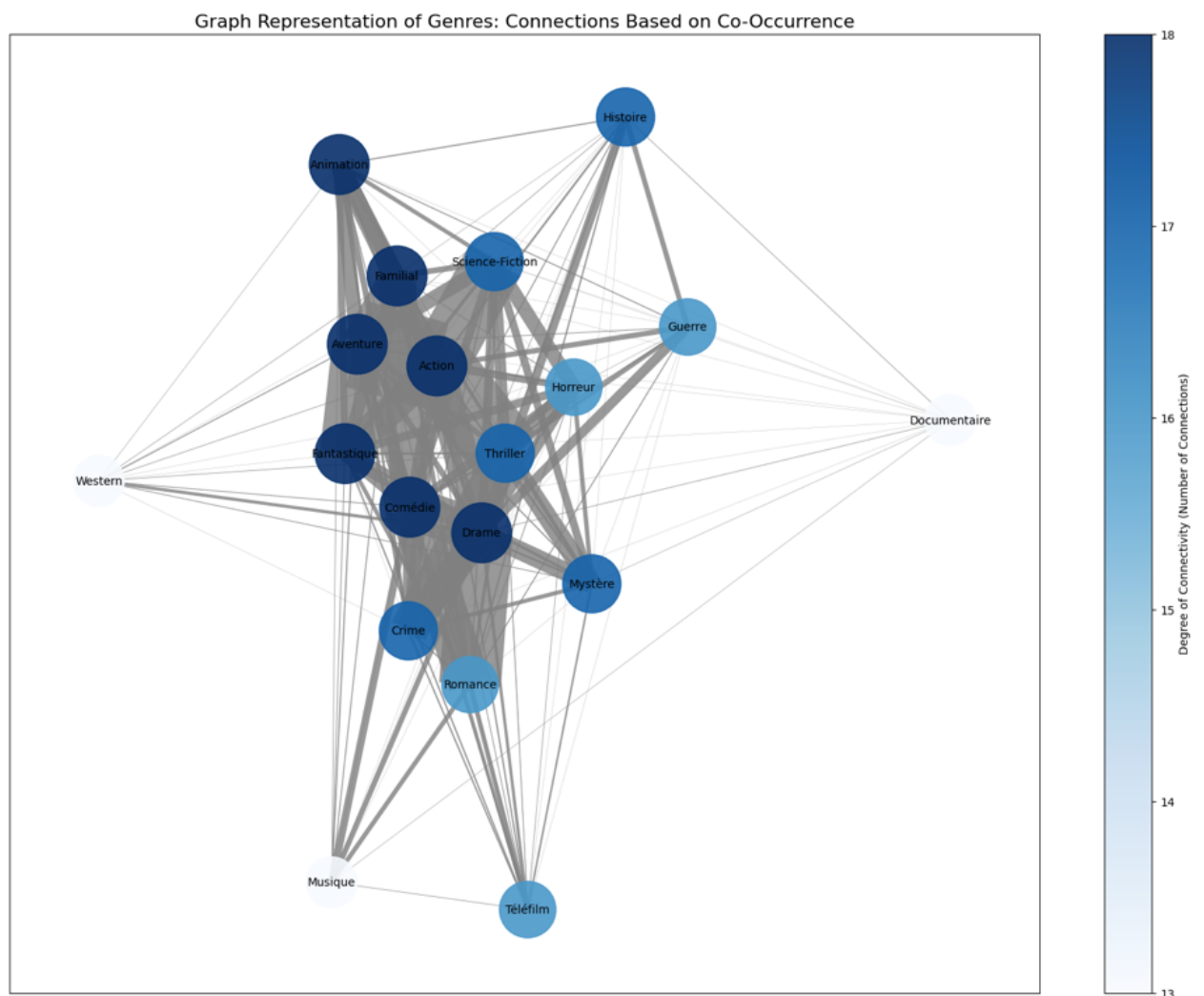
```

    edge_color="gray",
    alpha=0.8
)
nx.draw_networkx_labels(
    G_genres,
    pos,
    font_size=10,
    font_color="black"
)

# Ajouter une barre de couleur
plt.colorbar(nodes, label="Degree of Connectivity (Number of Connections)")

# Titre et affichage
plt.title("Graph Representation of Genres: Connections Based on Co-Occurrence")
plt.show()

```



## table de fréquence

```
In [17]: from collections import Counter
```

```

import pandas as pd

# Créer les relations entre genres
genre_edges = []

for _, row in pdf.iterrows():
    genres = row["Genres"].split(", ") if isinstance(row["Genres"], str)
    for i, genre1 in enumerate(genres):
        for genre2 in genres[i + 1:]:
            genre_edges.append((genre1, genre2))

# Compter les co-occurrences des paires de genres
edge_weights = Counter(genre_edges)

# Convertir en DataFrame pour une analyse facile
co_occurrence_df = pd.DataFrame(edge_weights.items(), columns=["Genre Pair", "Frequency"])
co_occurrence_df["Genre 1"] = co_occurrence_df["Genre Pair"].apply(lambda x: x.split()[0])
co_occurrence_df["Genre 2"] = co_occurrence_df["Genre Pair"].apply(lambda x: x.split()[1])
co_occurrence_df = co_occurrence_df.drop(columns=["Genre Pair"])

# Trier par fréquence décroissante
co_occurrence_df = co_occurrence_df.sort_values(by="Frequency", ascending=False)

# Afficher les 10 premières relations les plus fréquentes
print(co_occurrence_df.head(10))

```

	Frequency	Genre 1	Genre 2
47	384	Action	Aventure
35	314	Action	Thriller
24	268	Drame	Romance
10	259	Action	Science-Fiction
0	233	Animation	Familial
25	230	Drame	Thriller
43	217	Aventure	Fantastique
1	202	Animation	Comédie
44	192	Horreur	Thriller
83	187	Comédie	Drame

```

In [18]: # Compter combien de fois chaque genre apparaît
genre_centrality = Counter([genre for edge in genre_edges for genre in edge])

# Convertir en DataFrame
centrality_df = pd.DataFrame(genre_centrality.items(), columns=["Genre", "Degree"])
centrality_df = centrality_df.sort_values(by="Degree", ascending=False)

# Afficher les 10 genres les plus connectés
print(centrality_df.head(10))

```



	Genre	Degree
5	Action	2530
7	Drame	2453
8	Aventure	2122
2	Comédie	2075
9	Thriller	1903
3	Fantastique	1444
0	Animation	1376
1	Familial	1354
4	Science-Fiction	1260
6	Romance	1000

```
In [19]: # Filtrer les co-occurrences avec le genre "Action"
action_relations = co_occurrence_df[
    (co_occurrence_df["Genre 1"] == "Action") | (co_occurrence_df["Genre
]

# Trier par fréquence décroissante
action_relations = action_relations.sort_values(by="Frequency", ascending

# Afficher les genres les plus associés à "Action"
print(action_relations.head(10))
```

	Frequency	Genre 1	Genre 2
47	384	Action	Aventure
35	314	Action	Thriller
10	259	Action	Science-Fiction
13	164	Action	Crime
49	162	Action	Comédie
69	148	Action	Drame
54	130	Action	Fantastique
8	99	Aventure	Action
6	97	Science-Fiction	Action
63	81	Animation	Action