

Ma mission

Analyse de donnés par
Power BI

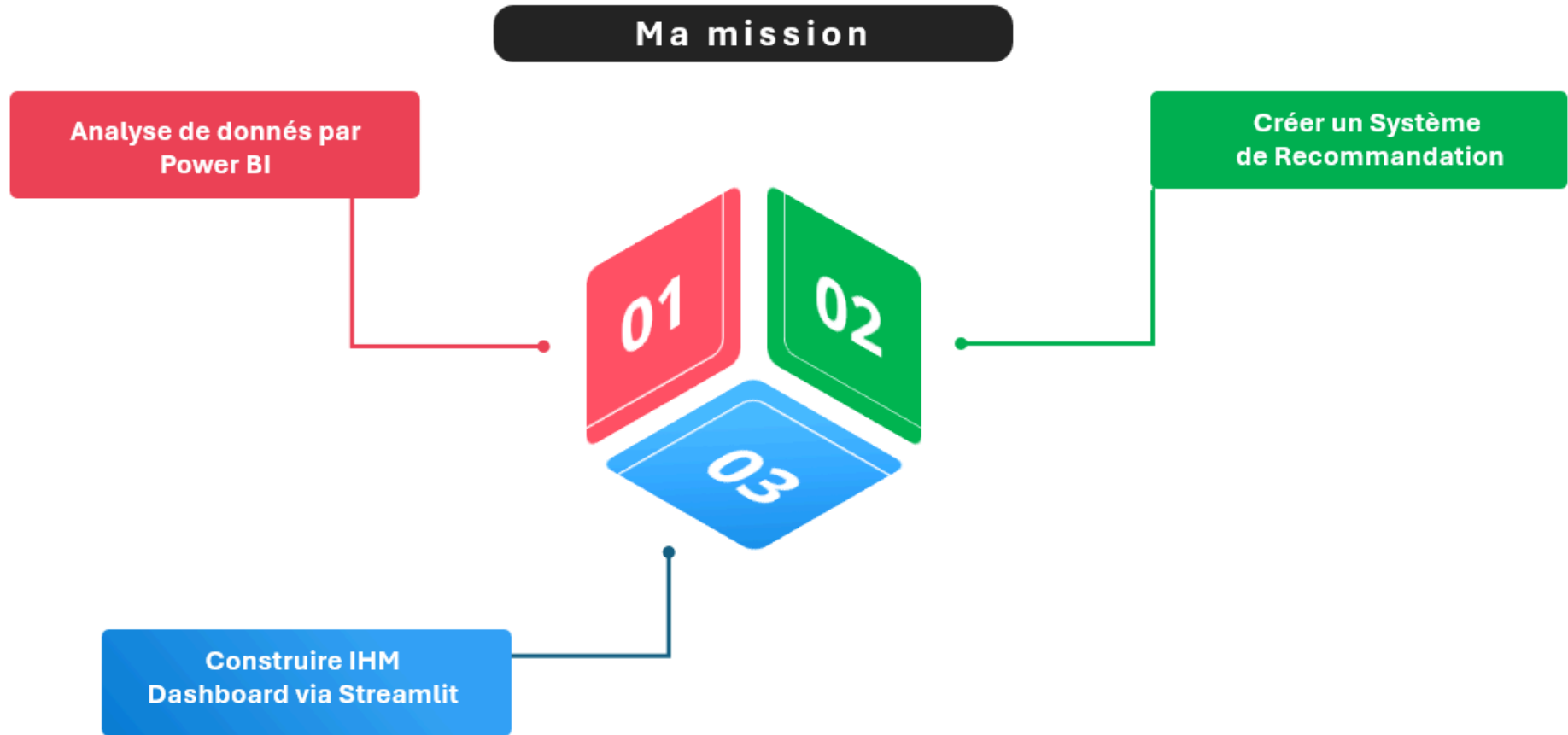
Créer un Système
de Recommandation

01

02

03

Construire IHM
Dashboard via Streamlit



Outils et Librairies



Plan

- ❑ **Présentation du jeu de données**
- ❑ **Analyse et visualisations via POWER BI**
- ❑ **Explication de l'approche de Système de Recommandation**
- ❑ **Présentation du Dashboard Streamlit**
- ❑ **Axe d'amélioration**

Présentation du jeu de données

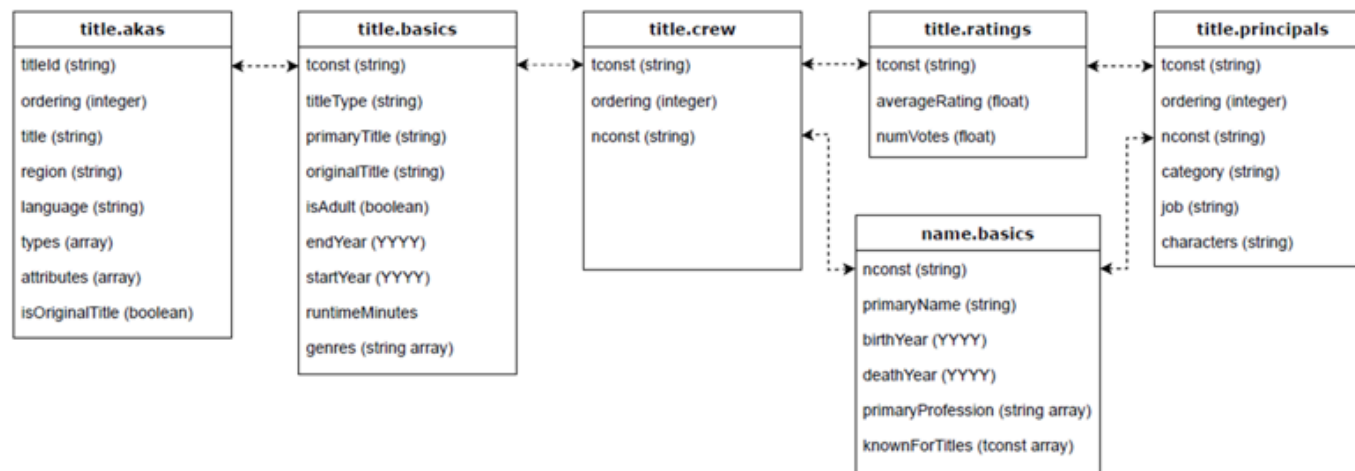
Pour cette mission, nous disposons deux sources de données principales :

- Base de données de IMDb composée de sept tables fournissant :

Informations principales sur les films
Details sur les contributeurs aux films
Evaluations des utilisateurs

- Base de données de TMDB contenant :

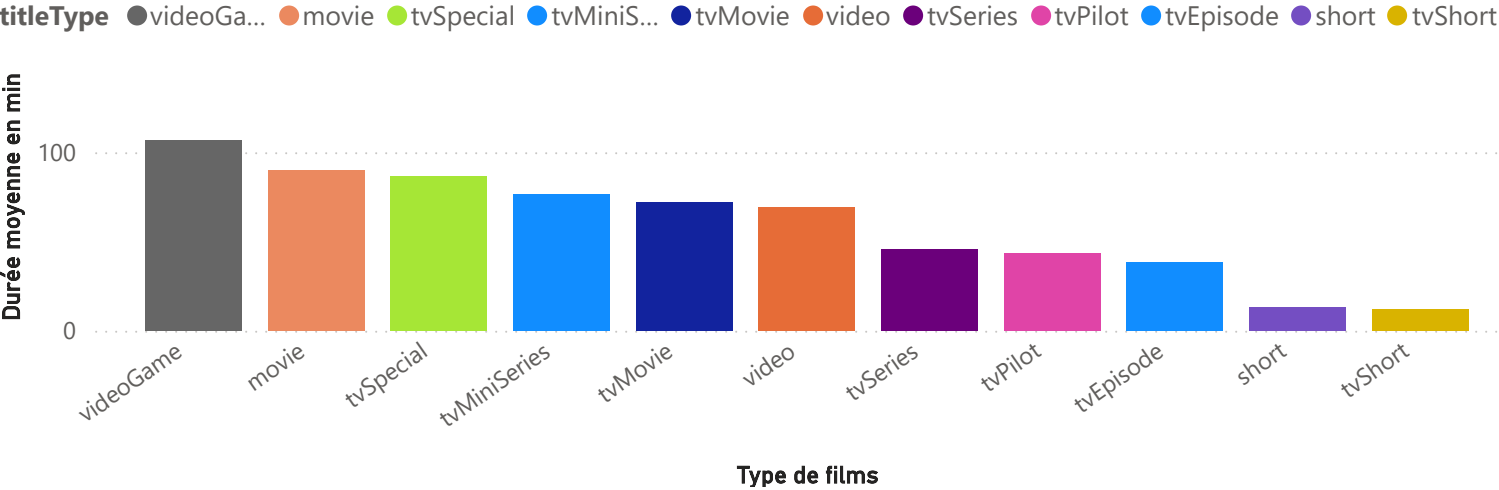
Informations sur les genres , budgets, popularité,
revenus des films



Nombre total
des films
tous types

10M

Durée moyenne par types de films présents



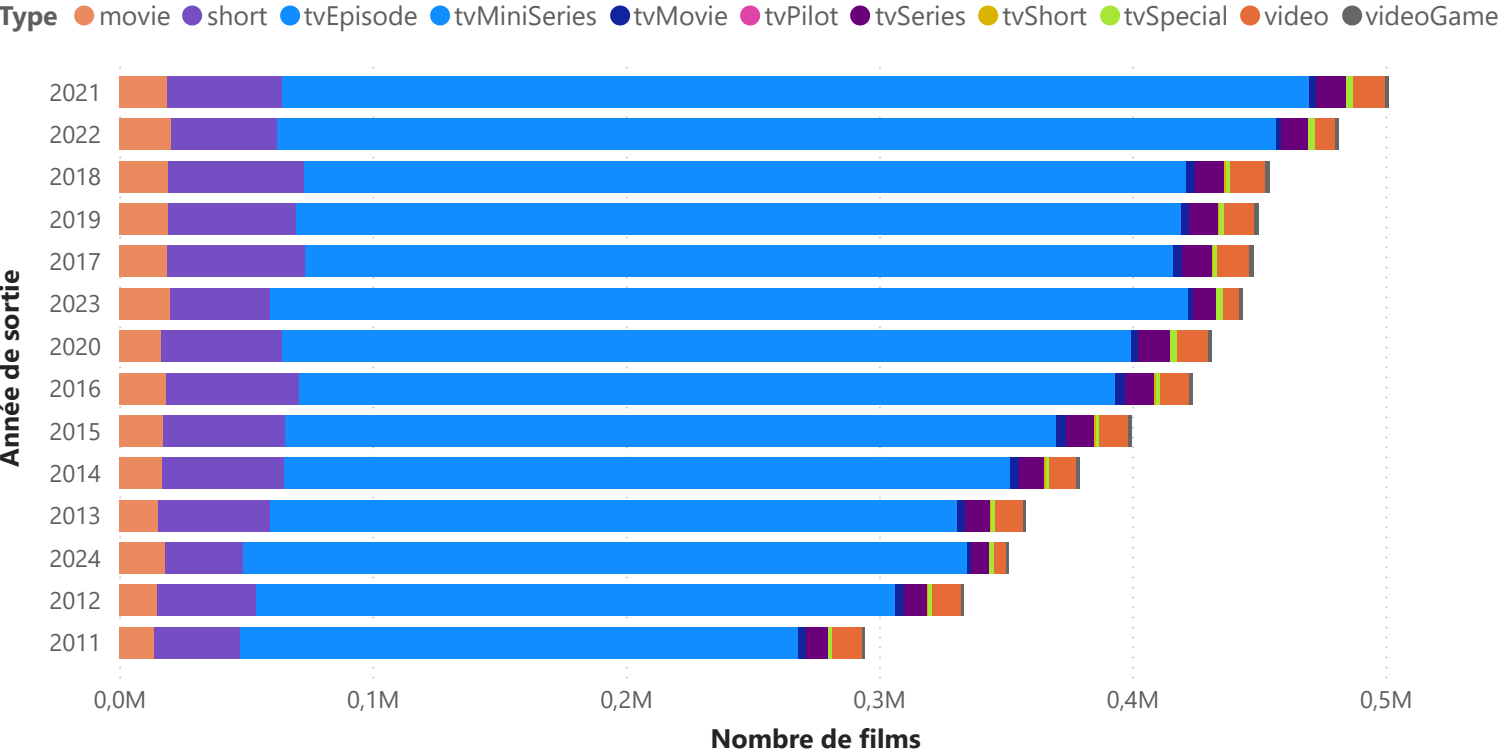
Année de sortie

Tout

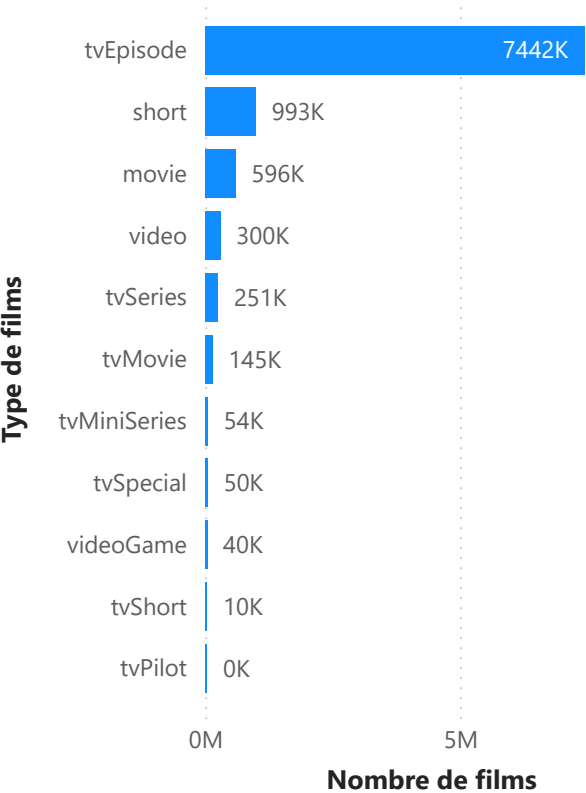
Type de films

- ☐ Sélectionner ...
- ☐ movie
- ☐ short
- ☐ tvEpisode
- ☐ tvMiniSeries
- ☐ tvMovie
- ☐ tvPilot
- ☐ tvSeries
- ☐ tvShort
- ☐ tvSpecial
- ☐ video
- ☐ videoGame

Répartition des titres de films par année et type



Nombre total de films par type



Focalisation l'analyse sur les 30 dernières années de production cinématographique pour garantir la pertinence de l'étude

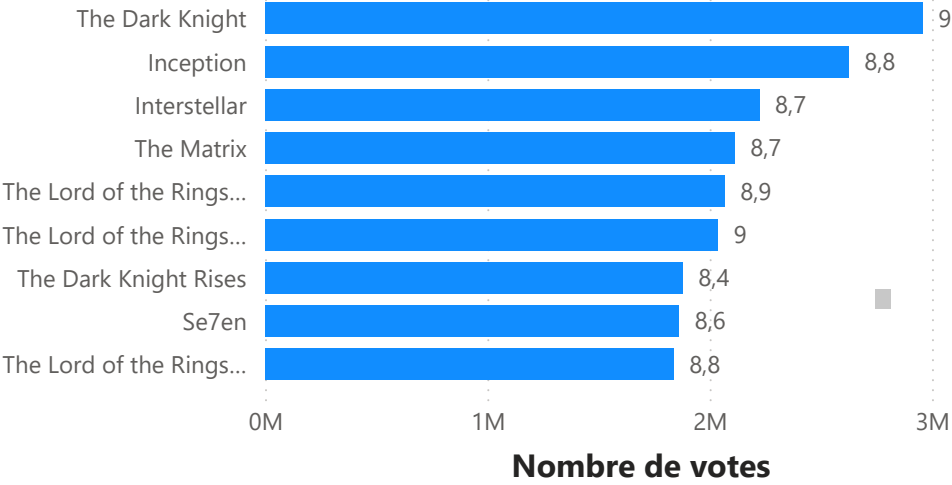
Année de sortie

Tout

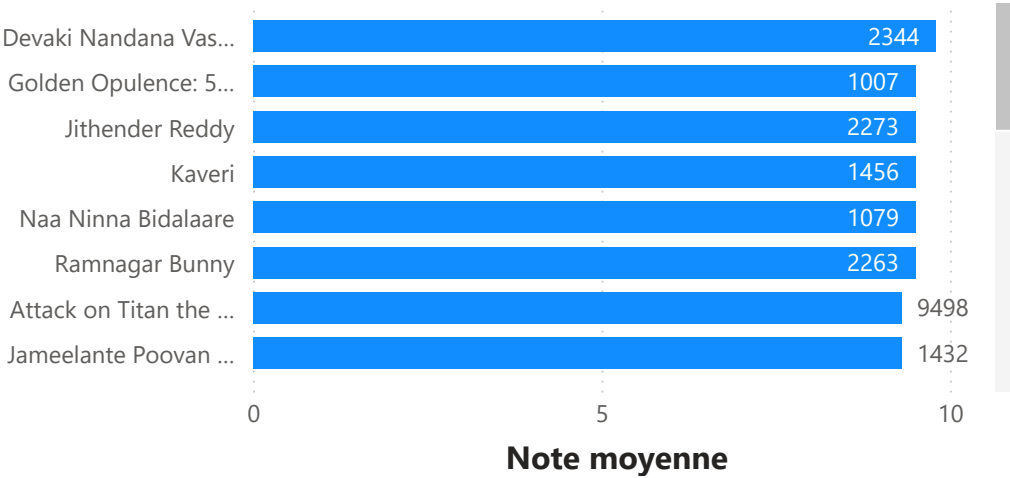
Nombre total
de movie

362K

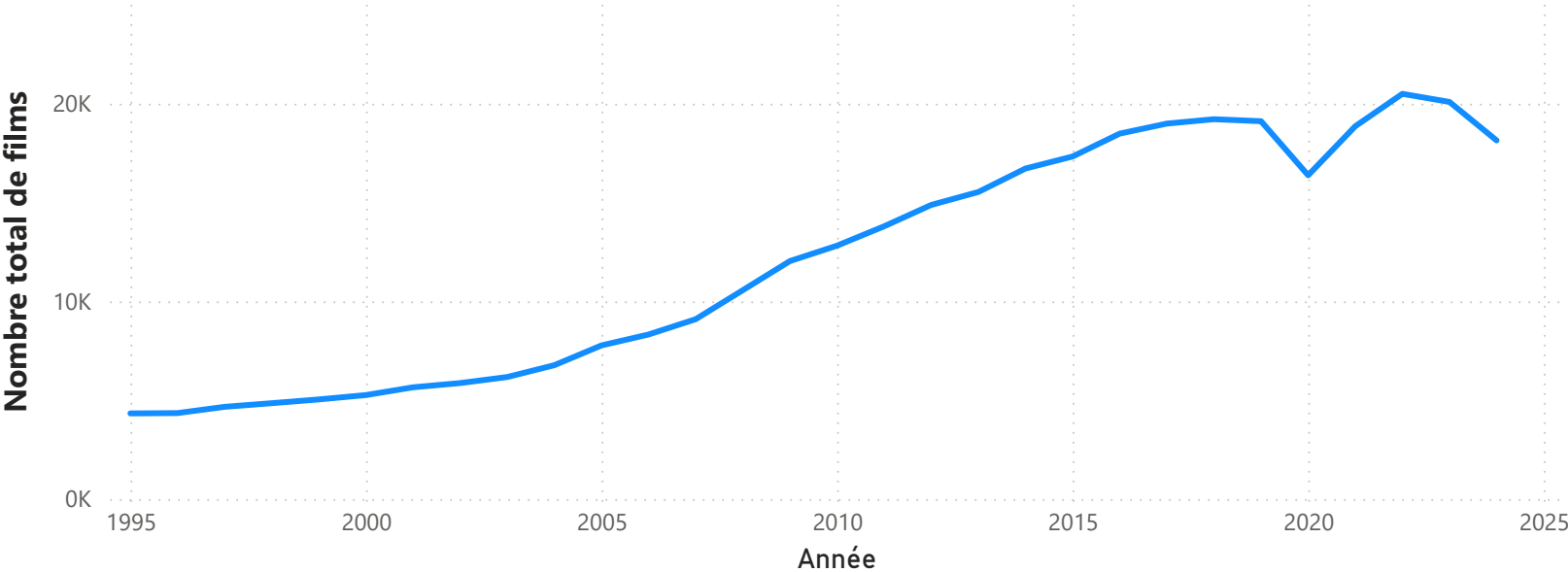
Top 20 des films ayant reçu le plus de votes



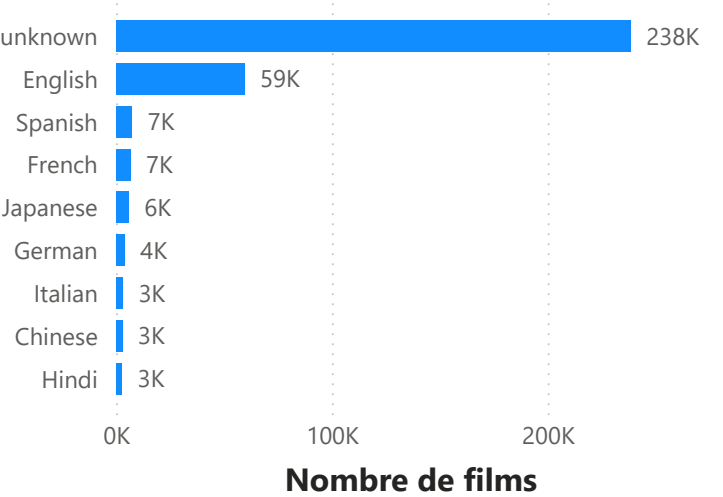
Top 20 Films avec les meilleures moyennes de notes (≥250 votes)



Évolution des pics de production cinématographique par année



Nombre de films par langue d'origine



Analyse des contributeurs les plus actifs et les mieux notés dans l'industrie cinématographique.

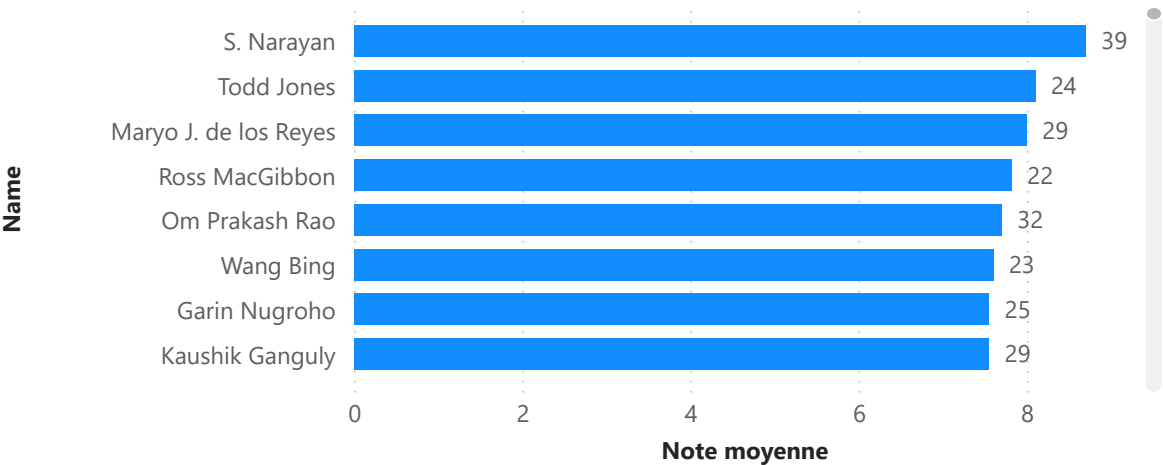
Ces graphiques montrant les écrivains, réalisateurs, actrices et acteurs ayant une note moyenne élevée, calculée uniquement sur des films ayant reçu au moins 500 votes , et pour lesquels chaque contributeur a participé à un nombre minimum de films.

Année de sortie

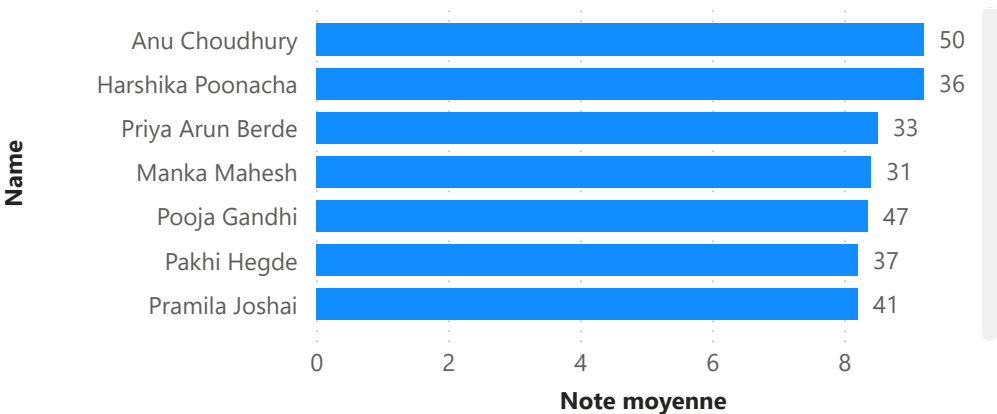
Tout



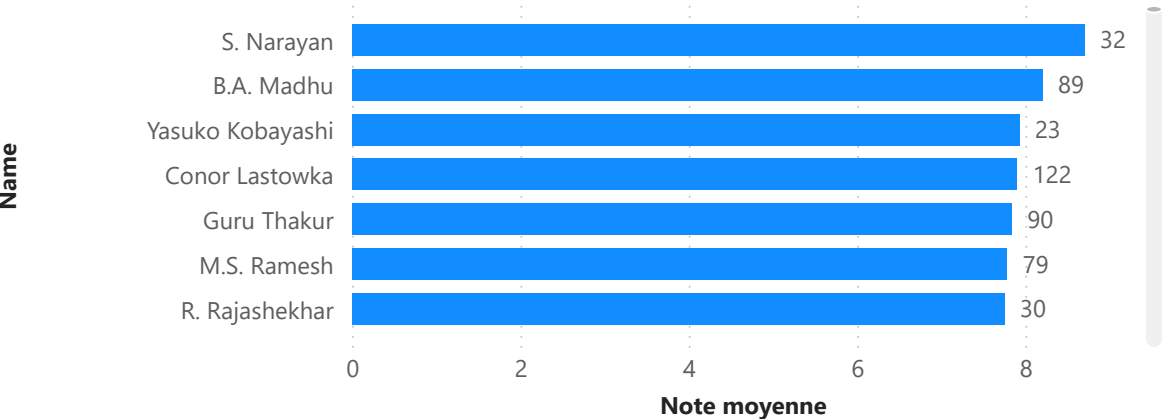
Réalisateurs les mieux notés



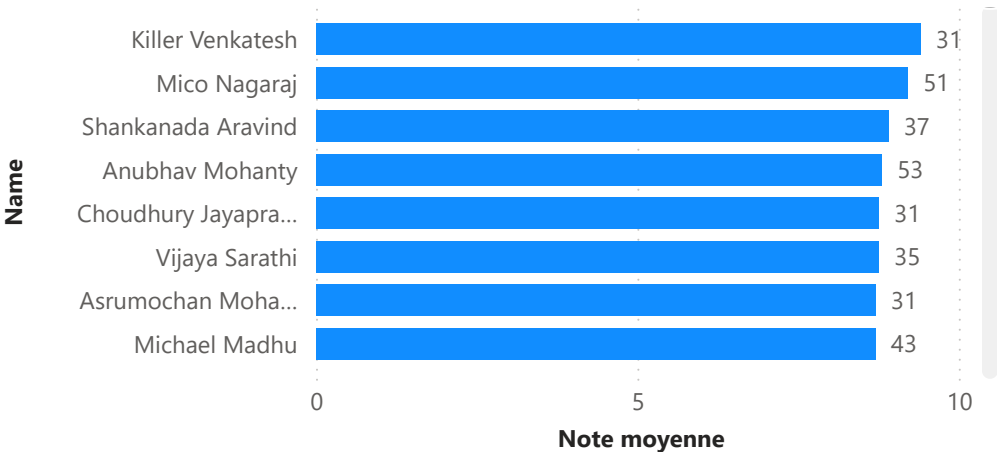
Actrices les mieux notées



Scénaristes les mieux notés

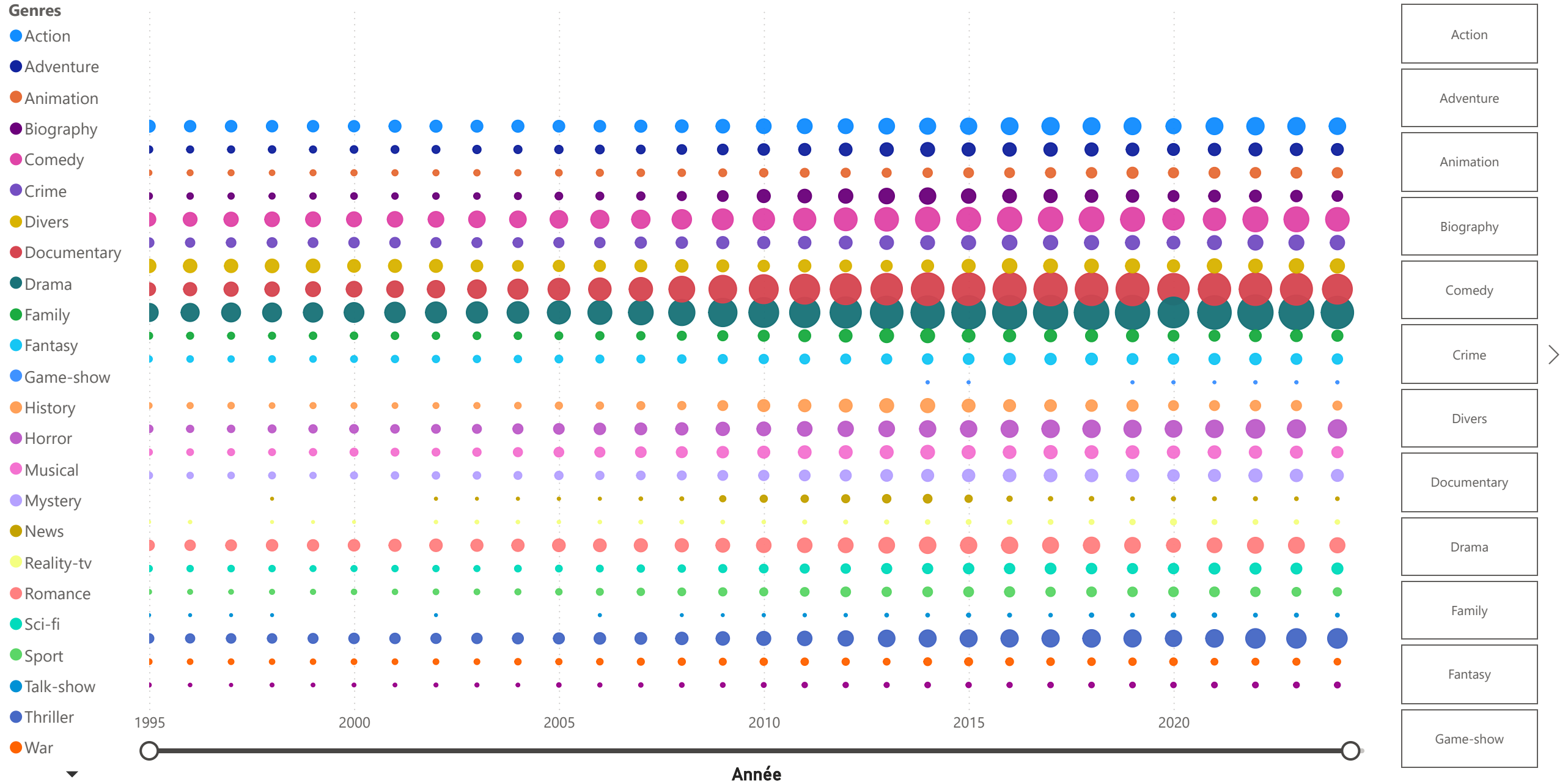


Acteurs les mieux notés



Évolution des genres de films par année

Ce graphique montre la répartition des genres de films au fil des années, où la taille des bulles indique le nombre de films produits pour chaque genre et chaque année.





Année

2024

Durée moyenne du movie (en min)

91.95

Nombre total de movie

18K

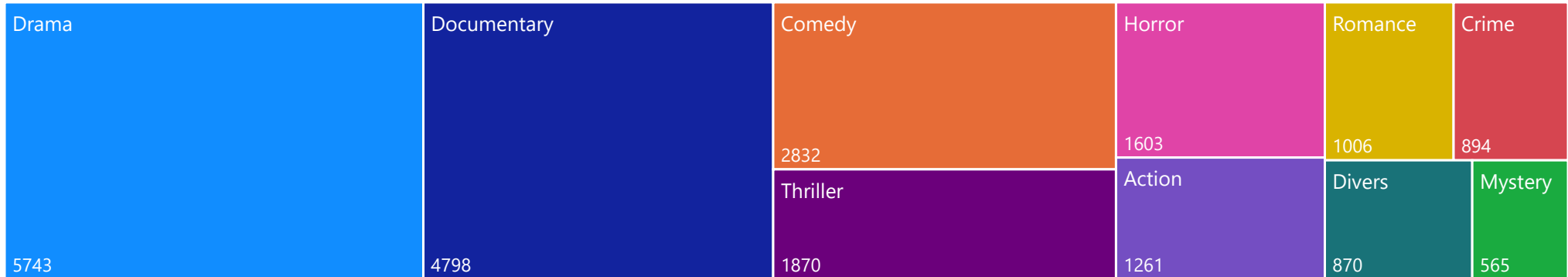
Nombre total de votes

13M

Note moyenne des films

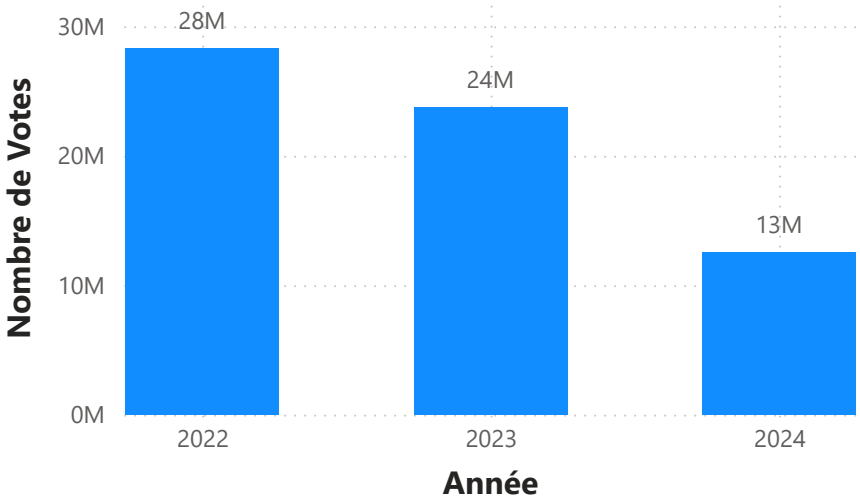
6,55

Genre le plus présenté

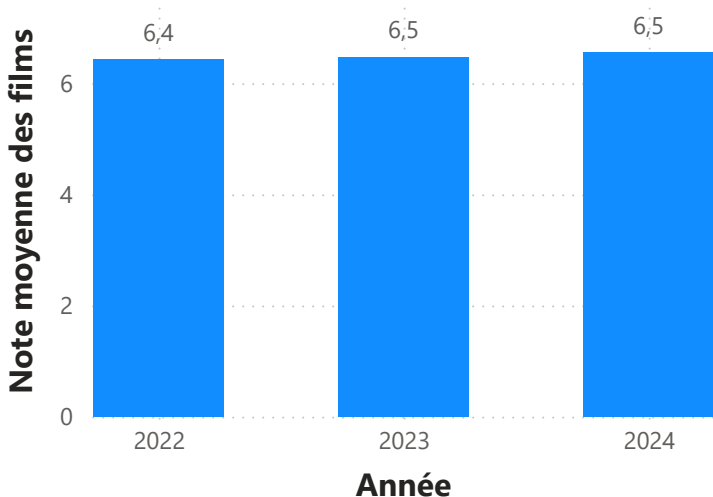


Analyse de l'évolution des métriques des films (2022-2023-2024)

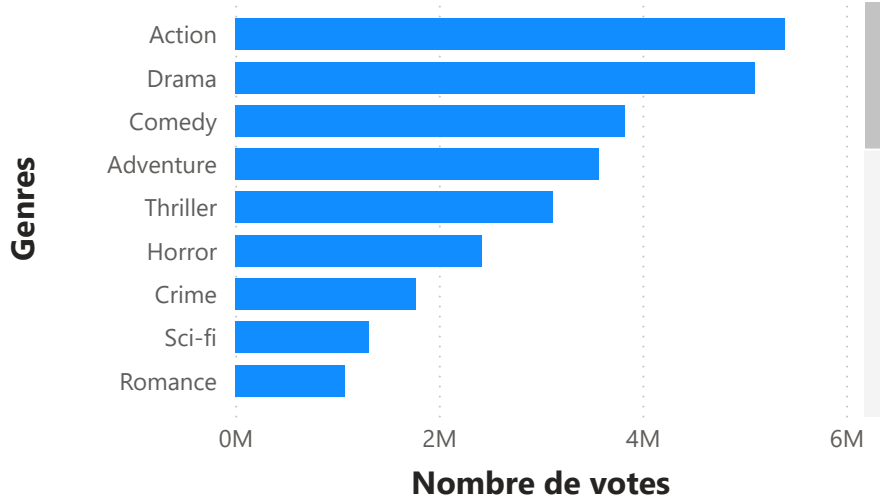
Total des votes par année



Évolution des notes moyennes des films



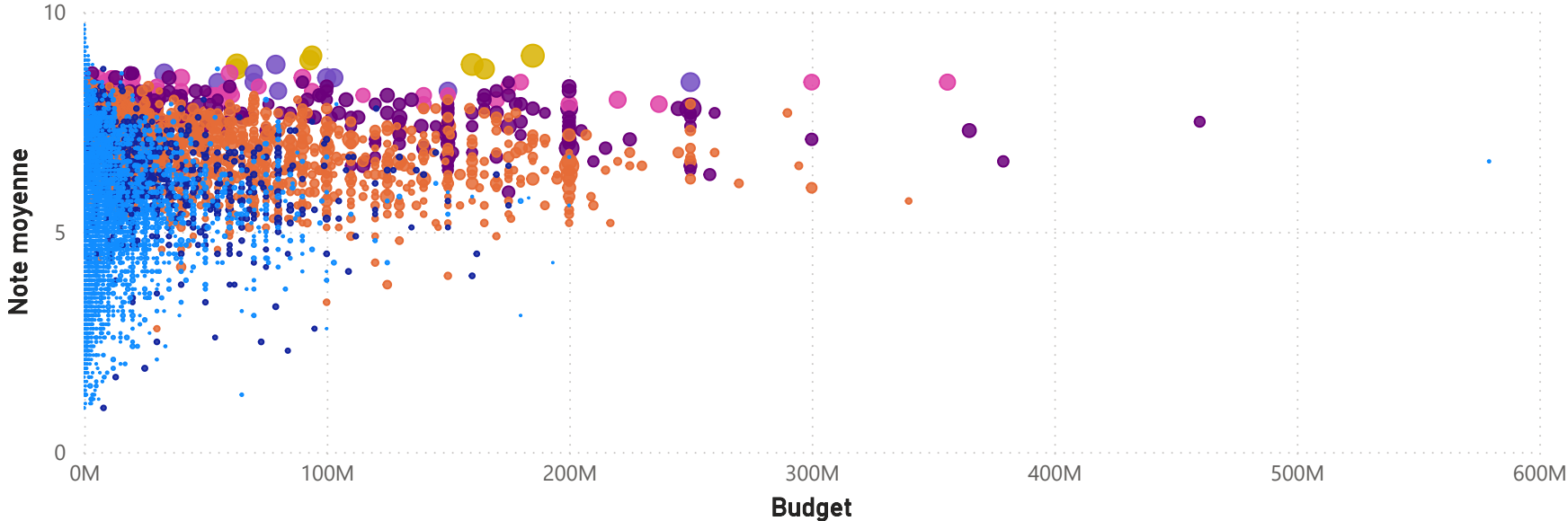
Analyse de la popularité des genres via les votes





Correlation entre le budget, la note moyenne et la popularité des films via le nombre de votes

CatégorieVotes < 50K 50K-100K 100K-500K 500K-1M 1M-1.5M 1.5M-2M > 2M



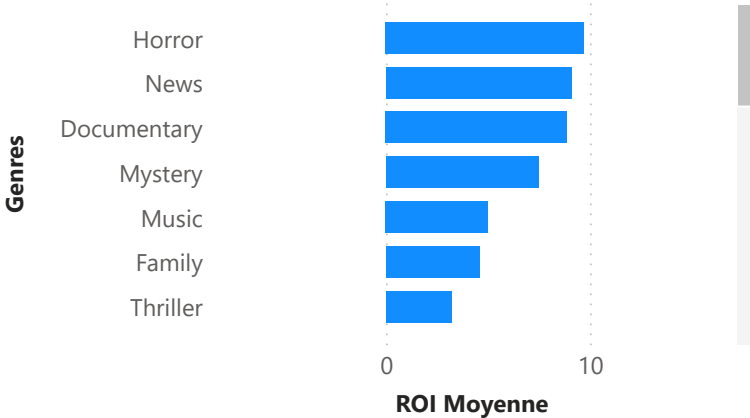
Année de sortie

1995

2024

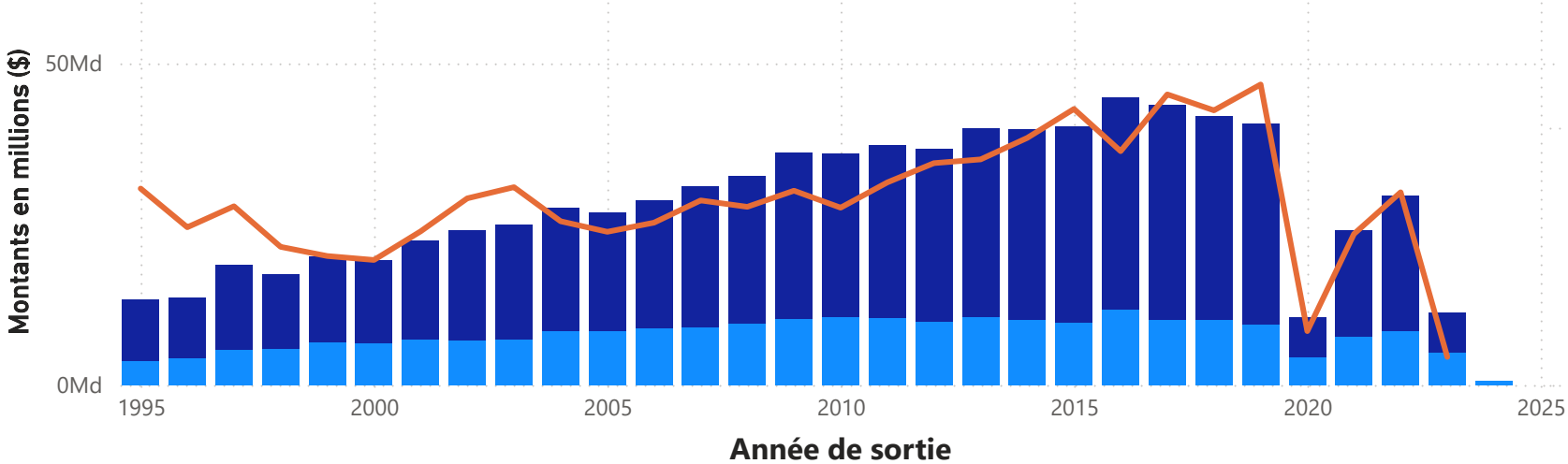
Genres plus lucratifs

ROI moyenne par genres



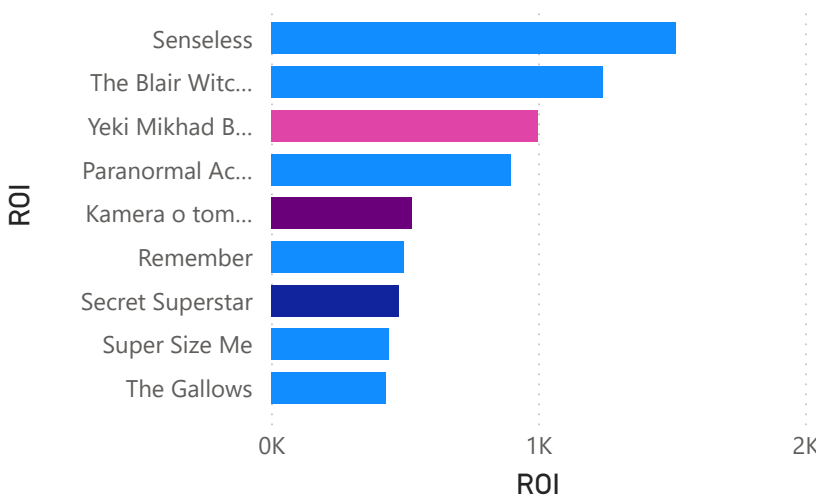
Évolution du Budget, des Revenus et du ROI par Année

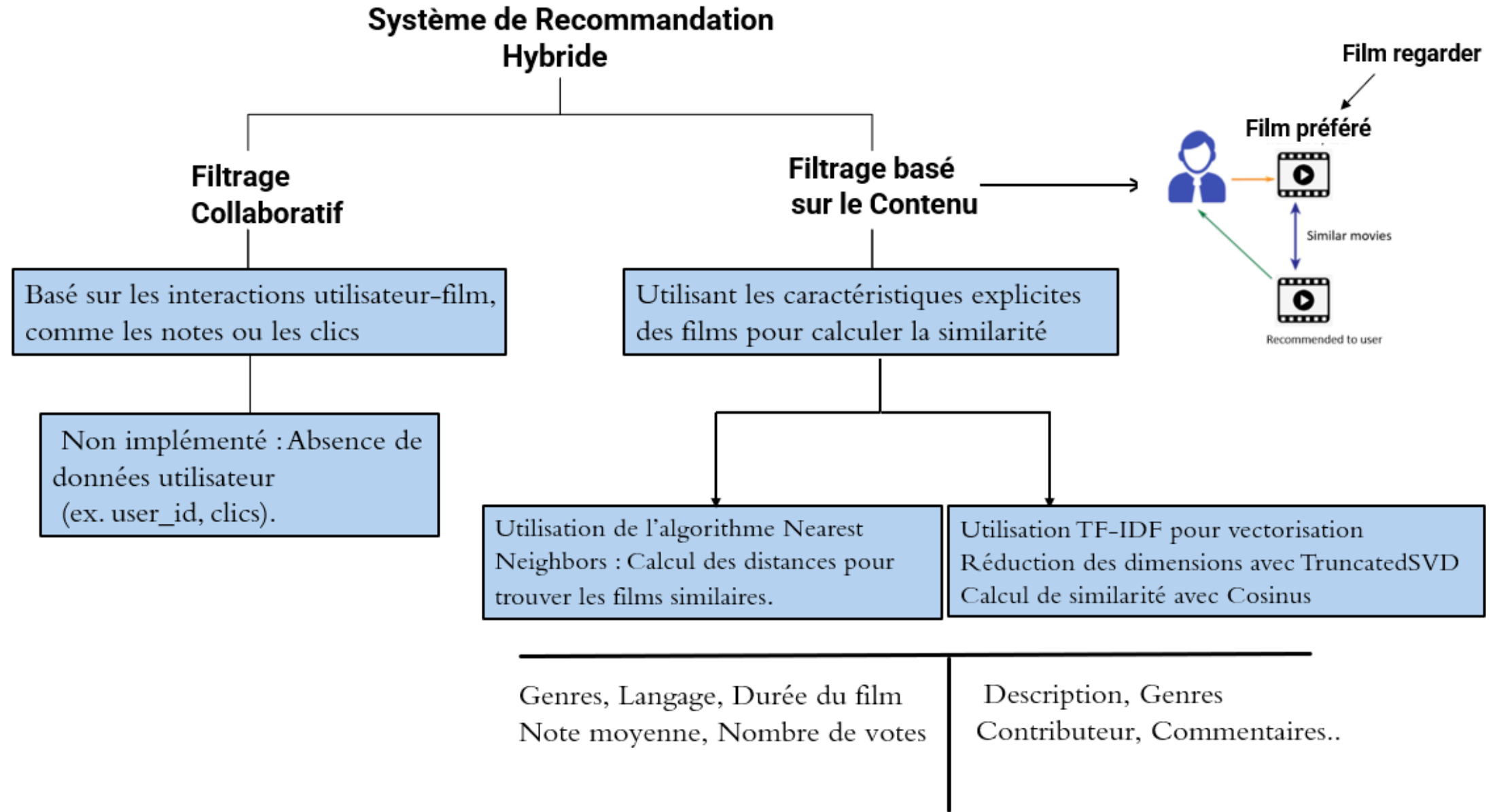
Somme de budget Somme de revenue ROI



Films les plus rentables

original_lang... English Hindi Italian Japanese





Data Preprocessing

01	Exploration de jeu de données et Identification des variables	<ul style="list-style-type: none">• Quantités de données et proportion de données manquantes• Quelles informations disponibles?
02	Sélections des variables	<ul style="list-style-type: none">• Identification des variables pertinentes pour l'étude• Suppression des colonnes non significatives
03	Analyse univariée	<ul style="list-style-type: none">• Résumé statistique avec describe() : moyenne, variance, médiane, min, max, quartiles• Visualisations countplot() pour catégories et distplot() pour les distributions numériques
04	Traitement des valeurs aberrantes et des doublons	<ul style="list-style-type: none">• détection des outliers sur la durée de film movie et suppression• Suppression des doublons dans les contributeurs
05	Traitement des valeurs manquantes	<ul style="list-style-type: none">• Suppression des colonnes avec plus de 80% de valeurs manquantes• Imputation des valeurs manquantes par la moyenne /médiane des films selon la catégorie
06	Encodage des variables	<ul style="list-style-type: none">• Application de OneHotEncoding pour les colonnes genres et langages
07	Standardisation	<ul style="list-style-type: none">• Application de StandarScaler uniquement sur les variables numériques (non encodes)

Présentation Dashboard Interactif

Fonctionnalité

- Construire un espace des caractéristiques pour vos films (genres, notes, votes, etc.)
- Prenez un film de référence (choisi par l'utilisateur)
- Choisir les préférences d'utilisateur : Genres / plage d'années de sortie/ acteurs spécifiques
- Utilisez NearestNeighbors pour trouver les K films les plus proches en termes de similarité.

Défis

- ☐ Proportion élevée de données manquantes
- ☐ Détection des outliers complexe
- ☐ Optimisation d'hyperparamètres du Modèle

Axe d'amélioration

- ☐ Déploiement du modèle sur le cloud
- ☐ Compte des utilisateurs
- ☐ Enrichissement des données