

DS53 : Analyse jeu de données 'The Movies Dataset' : aspects des dynamiques cinématographiques

Eléanore Renaud, Osman Gaygusuz, Gaspard Rochu

Sommaire

01

Contexte

Dataset utilisé, angle d'analyse,
modèle observé, etc.

02

Angle de vue numéro 1

Exploration des facteurs
déterminants du succès d'un film

03

Angle de vue numéro 2

Popularité des acteurs

04

Angle de vue numéro 3

Étude des préférences de genre
cinématographique

Contexte

- Source de donnée venant de deux datasets trouvé sur kaggle
 - <https://www.kaggle.com/datasets/akshaypawar7/millions-of-movies/data>
 - <https://www.kaggle.com/datasets/asaniczka/tmdb-movies-dataset-2023-930k-movies>

→ Données quotidiennement actualisé

- Modèle de donnée plats ou tabulaire:
 - Une seule table pour l'ensemble des données
 - Pas de relations complexes à gérer entre plusieurs tables.



Détails de la table

Détails des attributs :

1. **id**: An integer value serving as a unique identifier for each entry in the dataset.
2. **Title**:
3. **genres**: Categories or genres that the movie belongs to.
4. **Original_language**:
5. **overview**: A brief summary or description of the movie.
6. **popularity**: A float value indicating the popularity score of the movie.
7. **production_companies**: Names of the production companies involved in making the movie
8. **release_date**: The date when the movie was released.
9. **budget**: An integer value representing the budget of the movie.
10. **revenue**: An integer value representing the revenue generated by the movie.
11. **runtime**: The duration of the movie in minutes.
12. **Status**:
13. **tagline**: A short memorable phrase associated with the movie, often used in marketing.
14. **vote_average**: The average rating given to the movie.
15. **vote_count**: The count of votes given to the movie.
16. **Credits**:
17. **keywords**: Keywords or phrases associated with the movie for indexing and searching purposes
18. **Production_country** : pays de production du film (venant du deuxième dataset)



Contexte

Angle n°1

Angle n°2

Angle n°3

Traitement

Colonne personnalisée

Ajoutez une colonne calculée à partir des autres colonnes.

Nouveau nom de colonne

Profit

Formule de colonne personnalisée ⓘ

= [revenue]-[budget]

Colonnes disponibles

release_date
budget
revenue
runtime
status
tagline
vote_average

<< Insérer

En savoir plus sur les formules Power Query

- Utilisation de ~36 étapes de transformation de la donnée
 - Ajout de colonne (Profit,...)
 - Utilisation de Unpivot pour éviter d'avoir trop de colonne

	¹ ₃ id	^A _C title
1	615656	Meg 2: The Trench
2	615656	Meg 2: The Trench
3	615656	Meg 2: The Trench

^A _C Genre
Action
Science Fiction
Horror



Contexte

Angle n°1

Angle n°2

Angle n°3

Traitement

- Utilisation d'un script python pour traiter certaine donnée :
 - Le script traite les données de films à partir d'un fichier CSV.
 - Il agrège les données par acteur, calculant le nombre de films, le budget total, le revenu total et la moyenne des votes pour chaque acteur.
 - Les données agrégées sont sauvegardées dans un nouveau fichier CSV.

```
# Initialisation des dictionnaires pour stocker les données agrégées par acteur
films_count_by_actor = {}
total_budget_by_actor = {}
total_revenue_by_actor = {}
votes_by_actor = {}

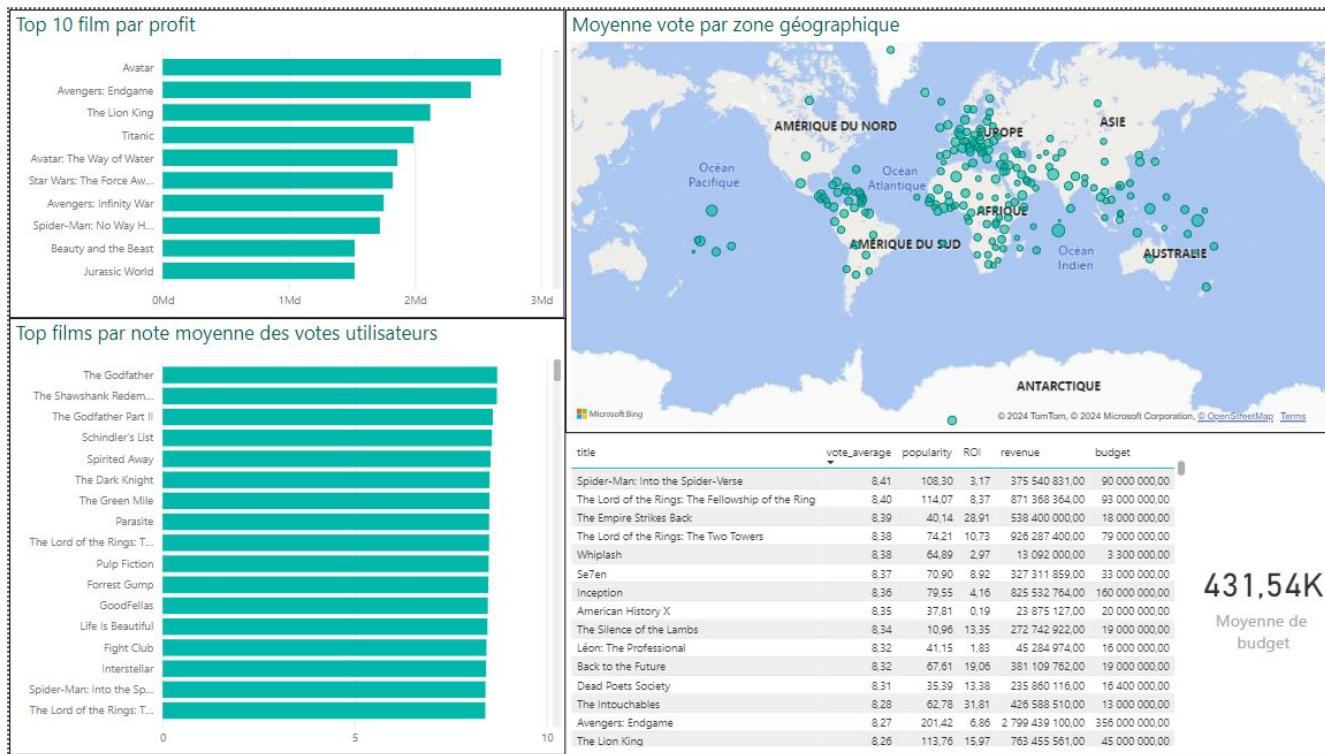
# Agrégation des données par acteur
for film in films_data:
    for actor in film['Actors']:
        if len(actor.split()) >= 2:
            votes_by_actor[actor] = votes_by_actor.get(actor, 0) + film['VoteA']
            # Nombre de films réalisés par chaque acteur
            films_count_by_actor[actor] = films_count_by_actor.get(actor, 0) + 1
            # Budget total des films dans lesquels chaque acteur a joué
            total_budget_by_actor[actor] = total_budget_by_actor.get(actor, 0) + film['Budget']
            # Revenu total généré par les films dans lesquels chaque acteur a joué
            total_revenue_by_actor[actor] = total_revenue_by_actor.get(actor, 0) + film['Revenue']
```

Angles de vue d'analyses

- **Success Metrics et Influences** : Exploration des facteurs déterminants du succès d'un film, en tenant compte de variables telles que la zone géographique, le budget, etc Films avec le meilleur retour sur investissement (ROI).
- **Popularité des acteurs** : Analysez la fréquence à laquelle certains acteurs apparaissent dans les films de votre dataset. Cela pourrait donner des indications sur les tendances de casting et les préférences du public.
- **Tendances des genres** : Étude des préférences de genre cinématographique pour dégager des tendances globales et régionales, tendances des genres de films populaires à travers les décennies.

Angle de vue N°1

Dashboard popularité films



Contexte

Angle n°1

Angle n°2

Angle n°3

8

Angle de vue N°1



- Exploration des facteurs déterminants du succès d'un film :
 - Zone géographique
 - Revenue, profit, budget etc.
 - ROI : Retour sur investissement
- Filtré avec un nombre de votes supérieur à 1000 et de vote moyen supérieur à 5
- Le score de popularité, établi par TMDB, est une mesure composite.



title	vote_average	popularity	ROI	revenue	budget
Spider-Man: Into the Spider-Verse	8,41	108,30	3,17	375 540 831,00	90 000 000,00
The Lord of the Rings: The Fellowship of the Ring	8,40	114,07	8,37	871 368 364,00	93 000 000,00
The Empire Strikes Back	8,39	40,14	28,91	538 400 000,00	18 000 000,00

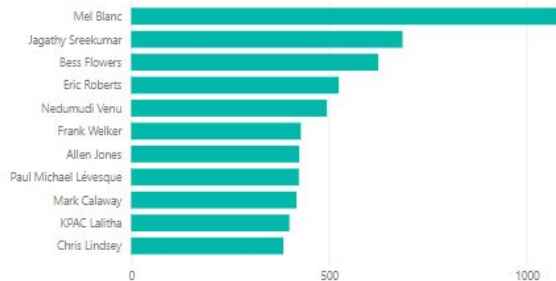
vote_average
est supérieur à 5

Type de filtre ⓘ

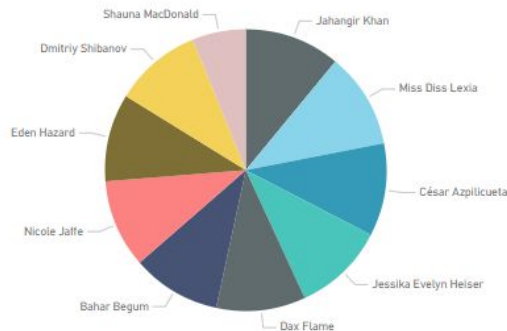
Angle de vue N°2

Dashboard popularité acteurs

Nombre de films dans lesquels des acteurs ont joué.



Notes moyenne des films joué par les acteurs

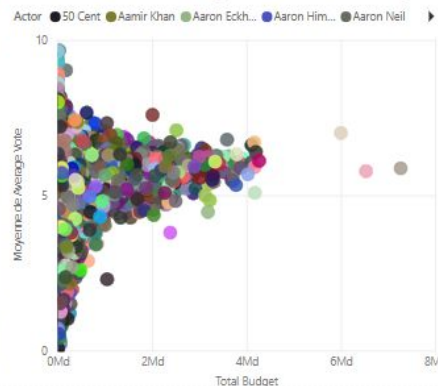


Budget totale et revenu total des films joué par acteurs



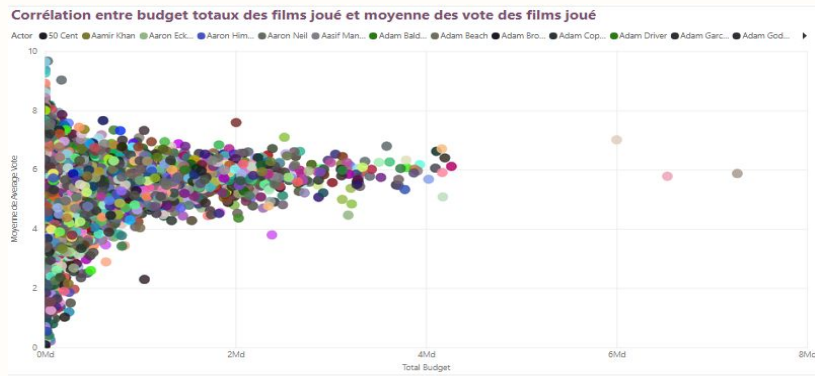
Actor	Films Count	Average Vote
Thales Egidio	6	9.97
Branislav Ivanović	11	9.94
Jack Schultze	3	9.93
Agha Mustafa Hassan	2	9.90
Gianfranco Zola	8	9.90
Kai Beavers	2	9.90
Phoebe Lovatt	2	9.90
Quinn Dunnett	2	9.90
Carlo Cudicini	7	9.89
Roberto Di Matteo	6	9.88
Gabriele Angleri	3	9.87
Graeme Le Saux	6	9.87
Jimmy Floyd Hasselbaink	6	9.87
Jody Morris	6	9.87
Mario Melchiot	6	9.87
Athmane Arioquet	5	9.86
Davi Claudino	5	9.86
Anjel White	4	9.85
Bijan Ganjali	2	9.85
Celestine Babayaro	5	9.84

Corrélation entre budget totaux des films joué par acteurs

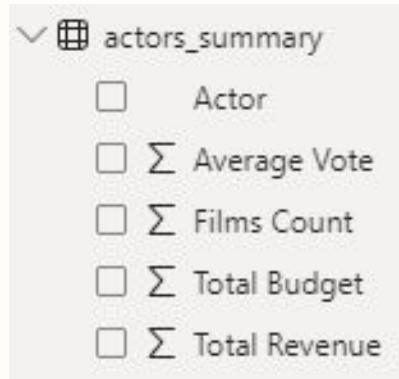
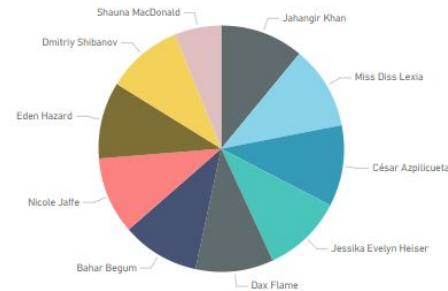


Angle de vue N°2

- Analyse de la fréquence d'apparition des acteurs dans les films : **tendances de casting** et **préférences du public**.
- Utilisation d'un script python pour changer les listes d'acteurs en table
- Filtrée avec un minimum de 15 film pour l'affichage en nuage de points



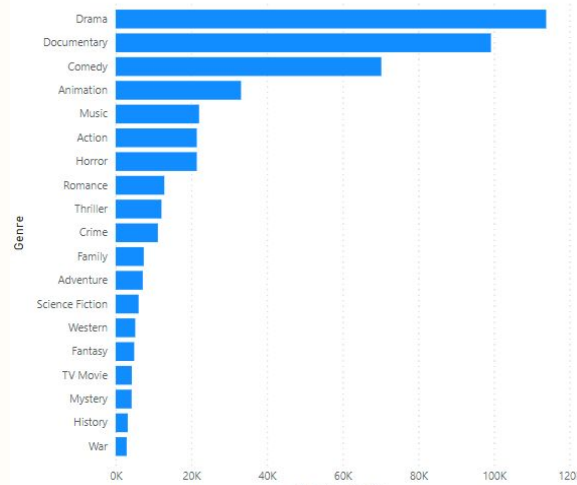
Notes moyenne des films joué par les acteurs



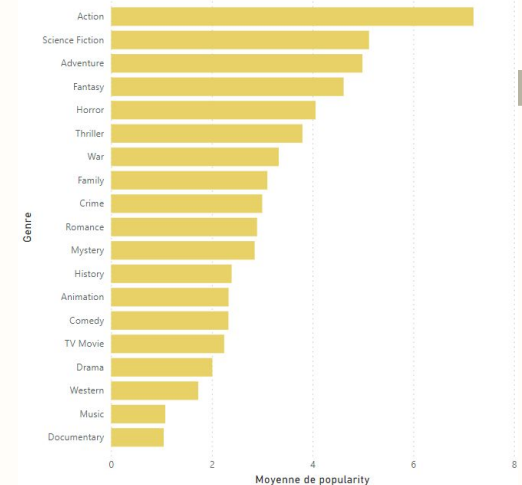
Angle de vue N°3

- Etude des préférences de genre cinématographiques
- Tendances des genres de films populaires et note moyenne donnée des films par genre
- Dégagement de corrélation entre budget alloué et genre d'un film

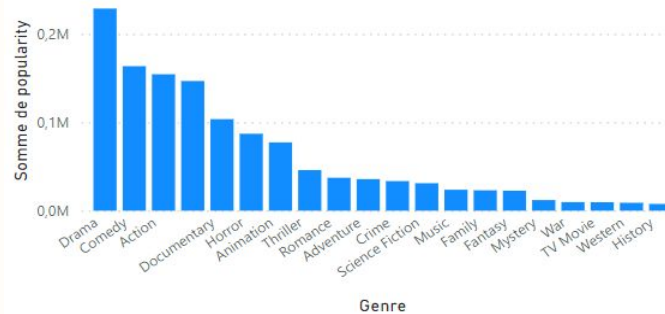
Nombre de film par Genre



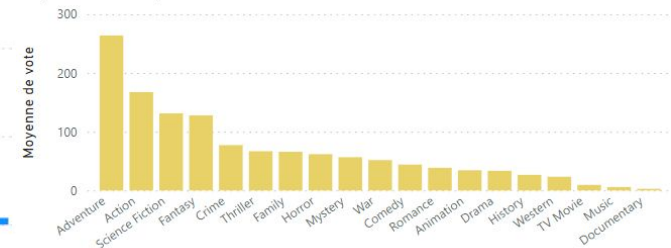
Moyenne de popularité par Genre



Somme de popularité par Genre



Moyenne de vote par Genre



662,08K

Nombre de Film

Contexte

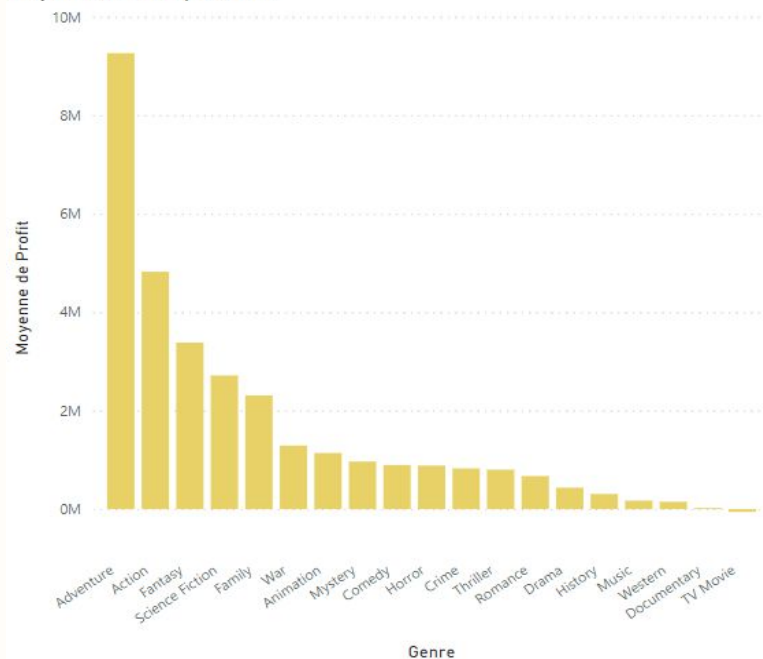
Angle n°1

Angle n°2

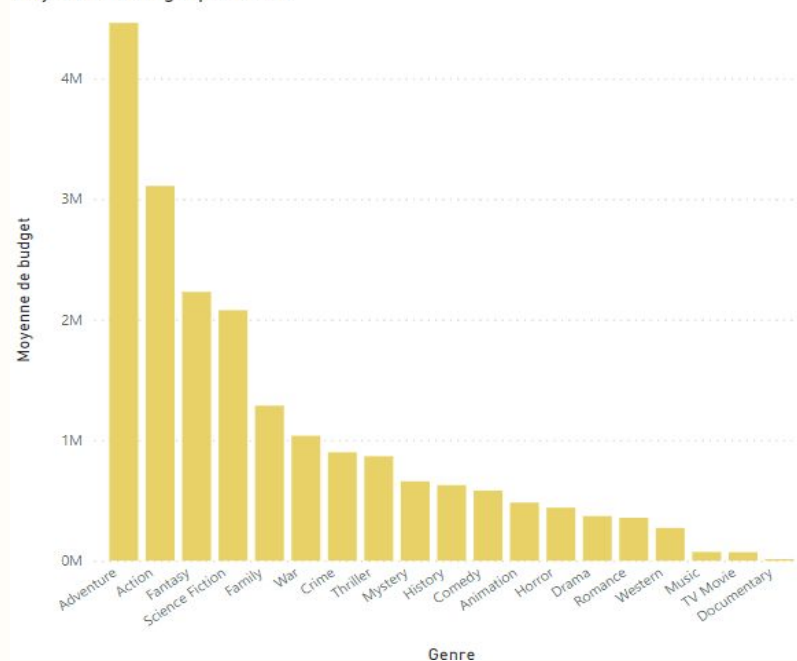
Angle n°3

Angle de vue N°3

Moyenne de Profit par Genre



Moyenne de budget par Genre



Démo

Conclusion



La popularité d'un film est principalement caractérisé par son budget, zone géographique, genre.



La popularité d'un acteur n'est pas caractérisé par le nombre de film joué



Le nombre de film par genre n'est pas témoins de popularité pour un genre

