

INSIGHT IN

2024

/

2025

VAISSADE – EL DAHDAH - TURPIN

MOVIES INDUSTRY

PAGE 02

SOMMAIRE

Dans ce projet, nous essaierons de donner une vue d’ensemble du segment marketing au sein de l’industrie cinématographique. En effet, ce secteur en plein essor cherche à rationaliser son fonctionnement et à le rendre plus transparent auprès des populations.

Pour y parvenir, nous présenterons d’abord les sources que vous avons retenues pour ce travail et les méthodes de traitement à l’origine de notre système d’informations. Dans un premier temps nous traiterons des contraintes liées aux sources individuelles et comment nous sommes parvenus à garantir l’intégrité des données dans le système et leurs fiabilités. Ensuite, nous proposerons une analyse à travers des graphiques clairs pour aider aux publics à comprendre les enjeux du secteur et aux acteurs du secteur de le transformer pour répondre aux aspirations des consommateurs.

PAGE 03

INTRODUCTION

**Problématique**

Quels sont les facteurs qui influencent le plus le classement des films ?

**Niveau de granularité**

Instantané périodique : on a

**Sources de données**

**Source 1 : Constantin**

* nom du film
* date de sortie
* description,
* langue d’origine,
* revenus généré (en $)
* nombre de vote,
* nb moyen de votes,
* note reçu, société de production,
* budget(en $)
* popularité (échelle 0 à 100)
* autre langue
* le genre
* la durée totale (min)
* le lien hypertexte de la page web du film

PAGE 04

MODELISATION DES SID

* Périodicité : La périodicité de mise à jour n’est pas spécie par contre la dernière mise à jour remonte 2017
* Téléchargement depuis Kaggle au format csv
* Les données de cette base semblent réutilisable sans réstricttion particulières

**Source 2 : Gabriel**

* + **Contenu des données**

**Titre du film** : Nom officiel du film.

**Date de sortie** : Date précise de sortie du film (AAAA-MM-JJ).

**Genre** : film par genre ("Drame", "Comédie", "Action", etc).

**Note moyenne** : Évaluation moyenne du film, basée sur des scores utilisateurs ou critiques (1 à 10).

**Score critique** : Score d'évaluation par les critiques (souvent un metascore ou une autre mesure).

**Acteurs principaux** : Liste des acteurs majeurs ou principaux pour chaque film.

**Réalisateur** : Nom du réalisateur.

**Scénariste(s)** : Nom(s) du ou des scénariste(s) ayant participé à l'écriture.

PAGE 05

SCHEMAS (PAPIER)

**Votes** : Nombre total de votes ou d’avis d’utilisateurs.

**La périodicité** de mise à jour et la dernière mise à jour : Pas spécifier

**Format des données** : CSV

**Accès** : Libre

**Droit de réutilisation des données :** les données sont libres de droit pour une réutilisation non commerciale.

**Source 3 : Paco**

* Liste des variables : Titre, Realisateurs, Date de sortie, Budget du film, Recette lors de la premiere semaine, recette du film en Amerique du nord, Recette dans les autres parties du globes, Recette dans le monde entier
* Périodicité : Film sortie entre 1986 et 2021
* Moyen d’accès aux données (API, téléchargement, autre… - *token* nécessaire ou non, libre ou non) : CSV
* Droit de ré-utilisation des données : Libr

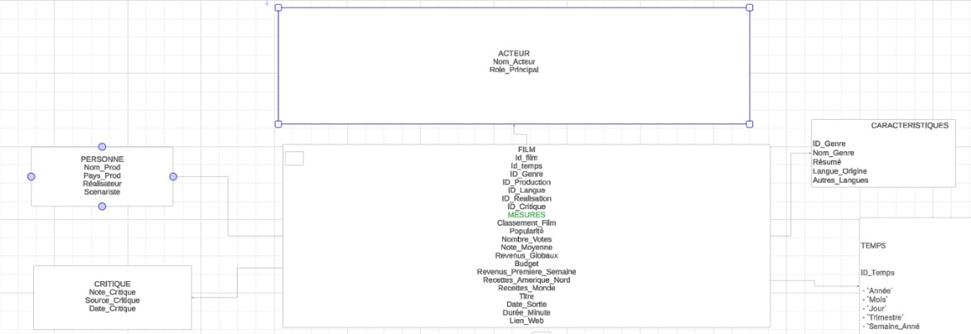
**Modèle logique des données**

****

PAGE 06

INTEGRATION DES DONNEES (1)

**Modèle en étoile centré sur le genre des films.**

****

Pour nettoyer la base de données j’ai d’abord supprimé les colonnes inutiles

j’ai remarqué qu’il y avait 3 lignes comportant des erreurs donc j’ai supprimé le diagnostic d’erreur et chargé la base dans power BI avec une nouvelle variable groupant les genres pour une meilleure interprétations futures des données.

PAGE 07

INTEGRATION DES DONNEES (2)

Ensuite j’ai réalisé un notebook python pour réaliser un tableau de bord sur les films

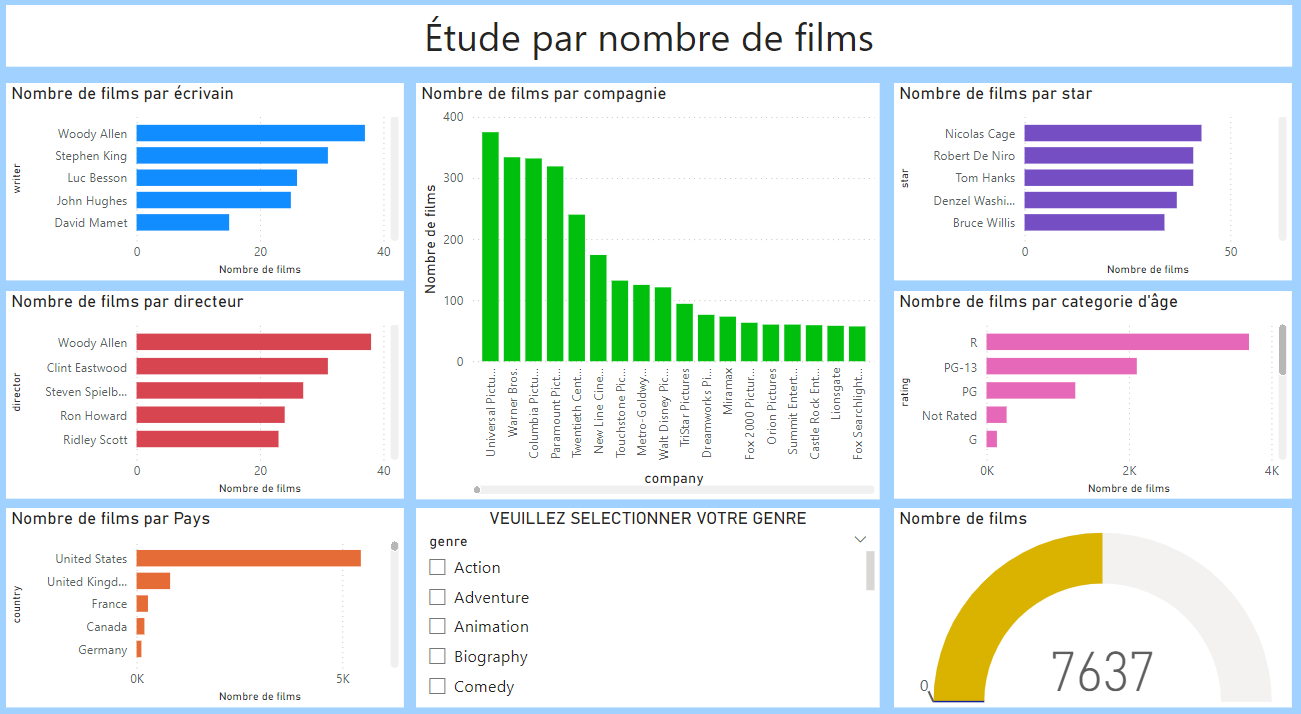
Voici un lien hypertexte pour y accéder :

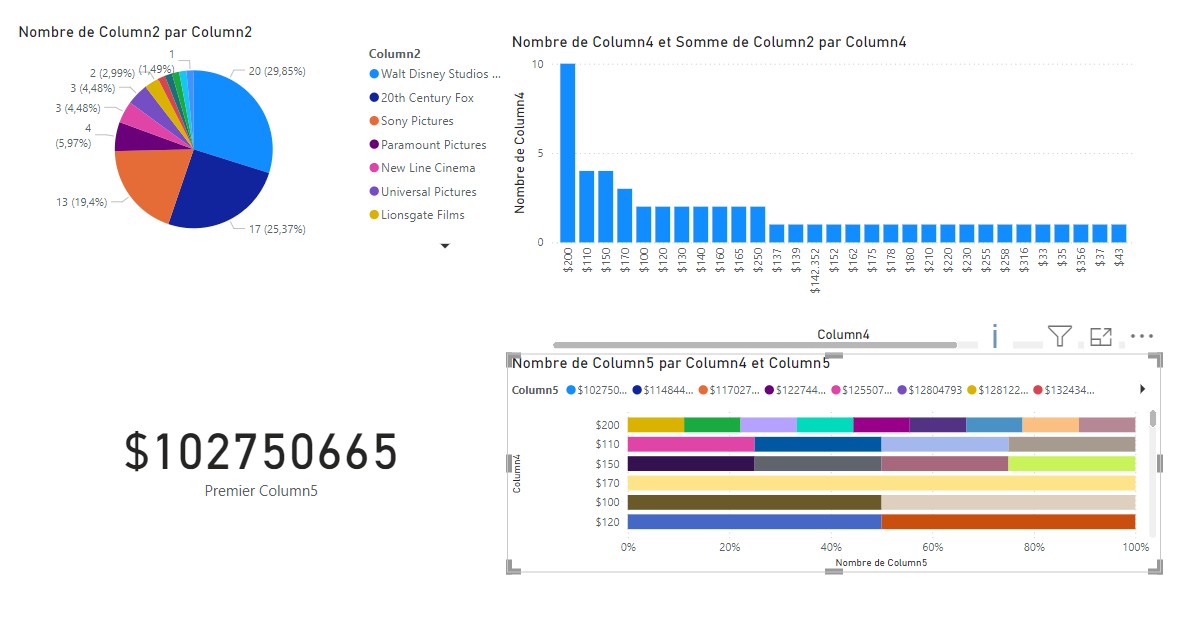
[**https://d.docs.live.net/1ee62e5815b5e32c/Bureau/BUT2/SID%20DateWarehouse/df.ipynb**](https://d.docs.live.net/1ee62e5815b5e32c/Bureau/BUT2/SID%20DateWarehouse/df.ipynb)

**Tableau de bord**

* **Classement par genre**
* **influence du casting sur le revenu géneré**
* **popularité par genre**
* **lien entre budget et revenus généré**
* **nombre de visite de la page du site film avant la sortie**
* **lien entre société de production et budget de réalisation**

**POWER BI pour la base de données “marvel”**

****

****

PAGE 09

RESTITUTION

Dans ce power bi il y’a 4 graphiques l’un représentant le pourcentage de film Marvel par studio de production. Le deuxième montre le nombre de colonnes ou le budget du film a coûté tant de dollars en millions. Le graphique en bas à gauche est une étiquette montrant un exemple des recettes d’entrée qu’a fait un film en Amérique du Nord (en l'occurrence dans le cas présent le film est Xmen : The Last Stand.

PAGE 10

CONCLUSION