Rapport projet SID

Constantin VAISSADE - Gabriel EL DAHDAH - Paco TURPIN

**Problématique**

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

**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
* 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.

**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.

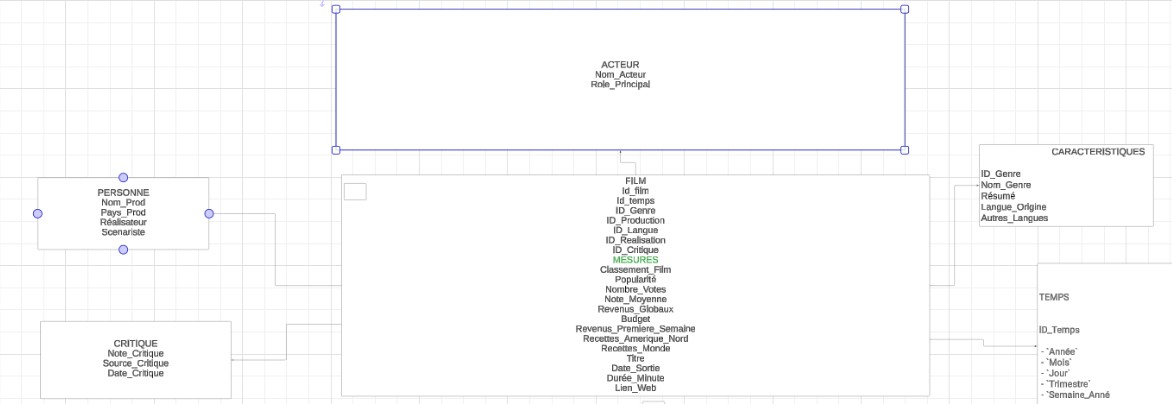
**Source 3 : Paco**

* 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 : Libre

Modele logique des données



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



<https://jupyter.org/try-jupyter/lab/>

Sur tmbd\_5000\_movie en CSV

Traitement via code python sur notebook consultable par le lien ci-dessus

Il y avait beaucoup de problèmes pour nettoyer la base de donner via le dax en effet pas

<https://d.docs.live.net/1ee62e5815b5e32c/Bureau/BUT2/SID%20DateWarehouse/df.ipynb>

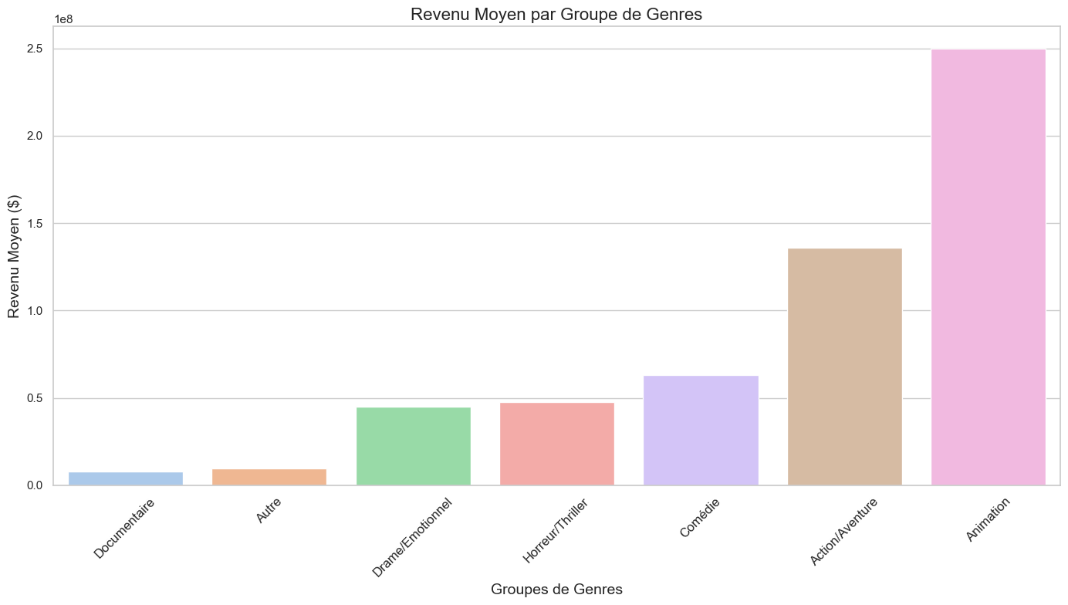
**Tableau de bord**

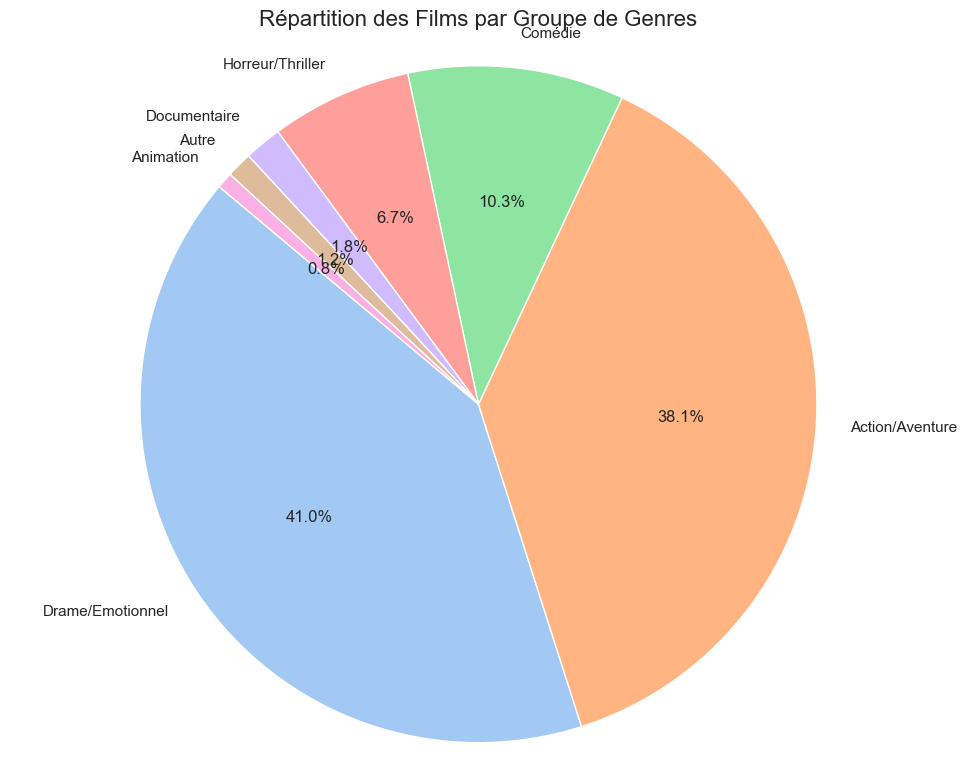
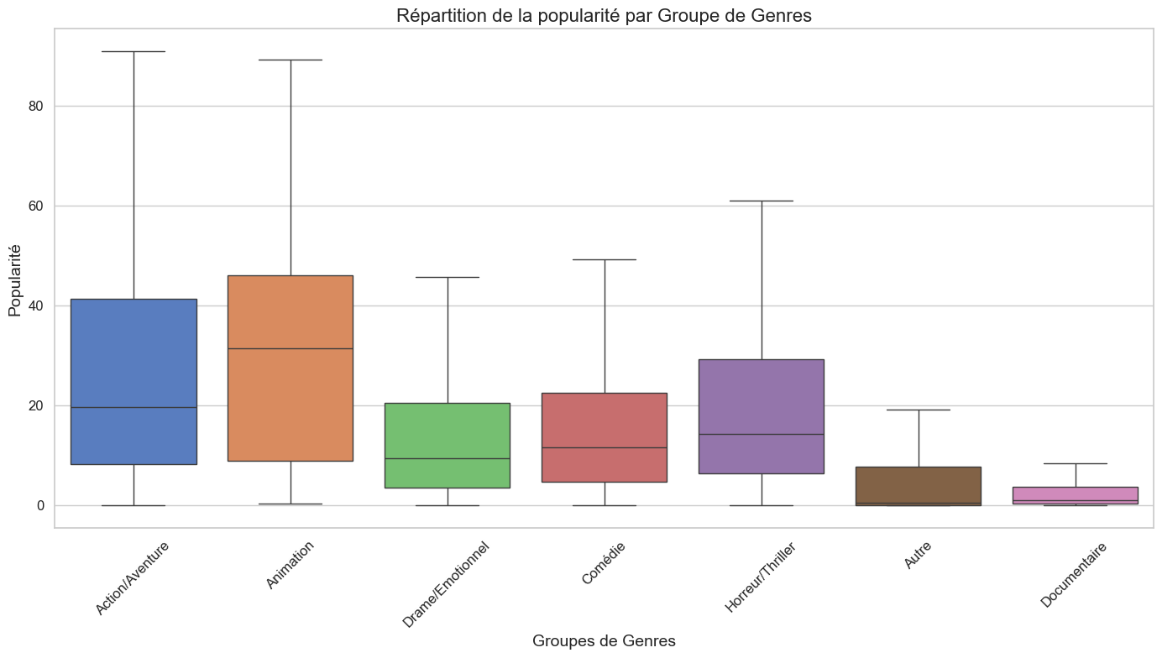
* **Classement par genre**
* **influence du casting sur le revenu géneré**
* **popularité par genre**
* **lien entre budget et revenus généré**
* **nombre de visite de la page du site film avant la sortie**
* **lien entre société de production et budget de réalisation**

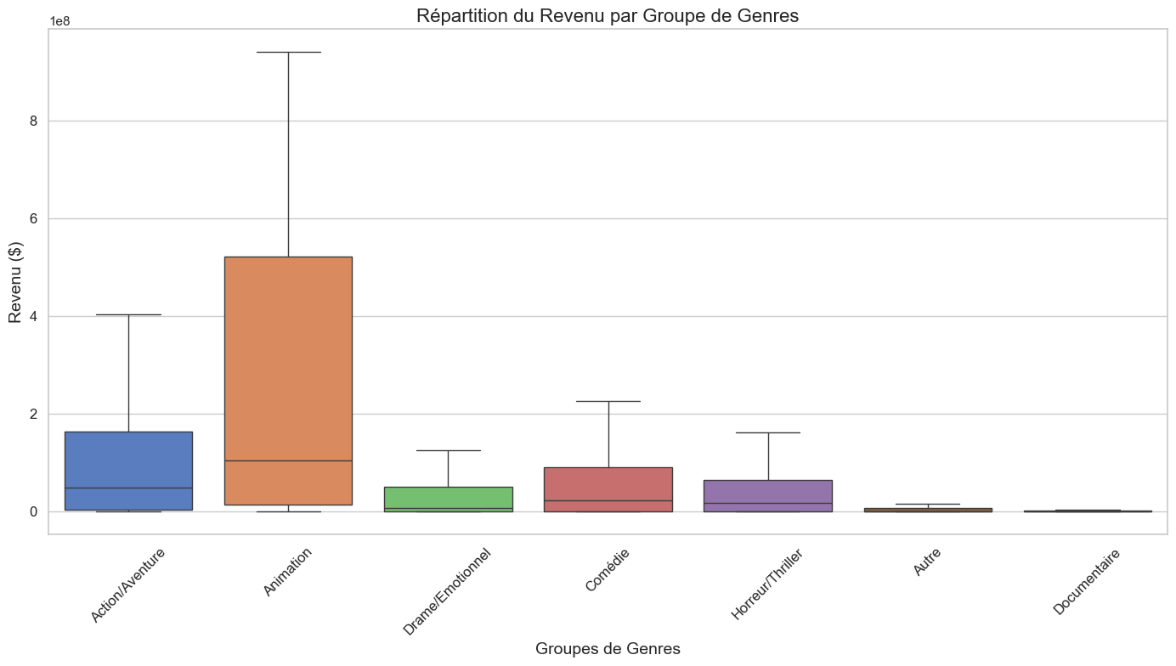
**onglet analyse genre par genre**

**revenue**

**répartition**

****

****

****