

Projet Analyse de Données

ROTH & BERTRAND & TREMBLAY & VALLANOVE & CHEVROT

4 novembre 2024

1 Contexte

Ce projet consiste à analyser une base de données.

Nous avons choisi une base de données contenant des informations sur des films populaires. Elle présente de nombreux films de différents genres, pays de production, budget, votes des spectateurs...

Nous allons utiliser cette base de données afin d'analyser les différents facteurs qui influencent le succès d'un film au box-office.

2 Structure de la base de données

La base de données contient 7668 lignes (et donc 7668 films) et est composée de 10 colonnes : name, rating, genre, year, score, votes, country, budget, gross et runtime.

3 Description des variables

- name (chaîne de caractères) : titre du film
- rating (chaîne de caractères) : classification qui décide si un film est approprié pour un public jeune ou s'il nécessite des restrictions en fonction de l'âge des spectateurs
- genre (chaîne de caractères) : genre du film
- year (entier) : année de sortie du film
- score (réel) : note moyenne du film allant de 0 à 10 donnée par les spectateurs
- votes (entier) : nombre de spectateurs ayant donné une note au film
- country (chaîne de caractères) : pays de production du film
- budget (entier) : budget de production du film en dollars
- gross (entier) : revenu total généré au box-office en dollars
- runtime (entier) : durée du film en minutes

4 Exemples d'utilité des variables

- rating : permet d'évaluer la tranche d'âge du public ciblé par le film
- genre : important pour analyser l'impact du genre sur les revenus (gross)
- year : permet d'observer des tendances dans le temps, notamment sur les budgets et recettes (budget et gross)

- score : peut d’indiquer la qualité du film
- votes : aide à mesurer la popularité d’un film
- country : peut être utile afin d’évaluer l’influence géographique sur les revenus (gross)
- budget : permet d’analyser si le succès d’un film est lié à son coût de production
- gross : variable essentielle pour mesurer le succès commercial du film
- runtime : pourrait influencer les spectateurs et donc le succès du film

5 Statistiques descriptives

Voici un résumé statistique des principales variables numériques de la base de données :

- score : La note moyenne des films est de 6.39 avec un écart-type de 0.97. La note la plus basse observée est 1.9 et la plus haute 9.3, ce qui montre une variation une variation significative de la qualité des films selon les spectateurs.
- votes : En moyenne chaque film reçoit environ 88109 votes avec un écart-type de 162925. Le nombre de votes varie de 7 à 2400000. La popularité des films présents dans la base est donc très variable.
- budget : Le budget moyen est d’environ 35.48 millions de dollars. L’écart-type est de 41.41 millions de dollars. Les budgets varient beaucoup, allant de 3000 dollars jusqu’à 356 millions de dollars.
- gross : Les films genèrent en moyenne des revenus de 78.5 millions de dollars avec un écart-type de 162.61 millions de dollars. La large plage de revenus, allant de 309 dollars à 2.85 milliards de dollars, montre les différences de succès commercial.
- runtime : La durée moyenne des films est de 107 minutes, avec un écart-type de 18.57 minutes. Les films durent entre 55 minutes et 366 minutes.

6 Conclusion et potentiel d’analyse

La base de données que nous avons choisie permet d’étudier les différents facteurs qui influencent le succès d’un film. Elle nous aidera à identifier les caractéristiques communes aux films les plus rentables, à analyser l’impact du budget ou de la durée sur les revenus, ou encore d’examiner les préférences des spectateurs en fonction du genre et du score par exemple.