

# Projet de visualisation STIC-B540)

Zaynab DOUDOUH, Bao TRAN, Hà Uyên TRAN

Novembre 2024

## 1 Introduction

Dans ce projet, nous avons exploré des ensembles de données portant sur des films, comprenant des informations telles que leur budget, leur popularité, leur genre, et leur pays de production. Les données proviennent de The Movie Database (TMDB) et incluent 23 variables principales, allant des revenus générés à la durée des films, en passant par leur langue originale et leur popularité. Notre objectif principal est d'identifier les facteurs qui influencent la rentabilité, l'appréciation auprès des internautes et la diversité de l'offre des films.

## 2 Méthodologie

Nous avons analysé la rentabilité des films (calculée comme la différence entre les revenus générés et le budget alloué) à travers plusieurs perspectives. Pour chaque question, nous avons utilisé des graphiques créés avec ggplot2 et avons justifié nos choix de représentations visuelles en tenant compte des particularités des données (valeurs extrêmes, distributions non uniformes).

### Q1: L'offre cinématographique

Dans cette question, nous allons nous intéresser à l'offre cinématographique. Cette notion est un concept large qui englobe plusieurs sous-concepts comme la diversité des genres, la variation des durées à travers le temps, etc. . .

#### Sorties par année

Pour commencer, nous allons nous intéresser à l'évaluation des sorties de films chaque année à partir du XIX<sup>ème</sup> siècle jusqu'à 2023 (2024 n'est pas terminé donc la base de données n'est pas complète). Nous pouvons observer dans le graphe que le nombre de sorties stagne entre 1930 et 1992 puis augmente considérablement. Cette augmentation est potentiellement due au rajout de film russe qui est inexistant entre 1916 et 1992.

Hypothèses: Ce vide dans les sorties peut être dû au fait que la base de données ne prennent en compte que les films sortis en Russie et non ceux issus de l'URSS, dont la Russie faisait partie pendant cette période.

## Genres de film produit par pays

Nous pouvons voir que la plupart des genres stagnent, à l'exception de quelques-uns. Les plus considérables sont les films documentaires, drama, actions et animations. L'augmentation des sorties du film, en plus du rajout de la Russie, est aussi causée par la croissance de ces genres. Nous remarquons qu'au XXI<sup>ème</sup> siècle, la Russie a sorti plus de films que les autres surtout pour de la drama, la comédie et l'animation. Le genre le plus produit au USA et au UK est les documentaires. Bien que les États-Unis produisent une grande quantité de films documentaires, leur nombre reste inférieur à celui de la Russie et du Royaume-Uni.

## Budget des films

Pourtant, même si la Russie produit plus de film que les États-Unis et le Royaume-Uni, son budget reste quand même inférieur. Nous pouvons le voir dans ce graphe: Chaque point représente un film et la taille représente son budget. Nous pouvons voir qu'il y a plus de gros points (donc de film avec gros budget) aux USA et au UK qu'en Russie qui a aucun film dépassant le 200.000.000 de dollar américain. Le leader en film à gros budget est les américains. Beaucoup de films produits dépassant les 200 millions de budgets sont généralement des films d'actions et d'aventures.

## Durée des films

Concernant la durée des films produits, nous pouvons voir dans ce graphique qu'il y a le plus de films vers 90 min. Le second pic de film se trouve vers 5 min.

Remarque: Pour les deux prochains graphiques, nous avons enlevé les films ayant une durée de plus de 250 minutes. La suppression de ces films est due au fait que ces films sont vraiment très peu et influencent inutilement l'échelle des graphes. (à reformuler)

Hypothèses: Beaucoup de film à 5min signifie qu'il y a beaucoup de court-métrage dans la base de données.

Pour terminer, voici une carte thermique représentant la durée moyenne de chaque genre. Nous pouvons observer que les films d'animations, peu importe le pays, sont généralement courts. Plus la couleur est claire, plus la moyenne des durées de film est grande. Nous pouvons comparer les différents de durées entre chaque pays. Par exemple, les films russes sont généralement plus long que les autres (il y a plus de rectangle clair). Contrairement à la Russie, le royaume uni ont peu voir pas de rectangle de couleur rouge clair. La moyenne des films britanniques sont plus court.

## **Conclusion**

### **Q2: Facteurs d'appréciation**

**Subset utilisé**

**Genre des films**

**Popularité des films**

### **Q3: Facteurs de profitabilité**

#### **Relation entre le budget et le revenu**

Objectif : Comprendre si un budget élevé conduit à un revenu plus important. Graphique utilisé : Nuage de points avec ligne de régression.

Ce graphique met en évidence une tendance générale où les films avec un budget élevé génèrent souvent des revenus plus importants. Cependant, la forte dispersion des points montre qu'un budget élevé ne garantit pas toujours le succès financier. L'utilisation de l'échelle logarithmique améliore la lisibilité des petites valeurs.

#### **Profitabilité en fonction des genres**

Objectif : Identifier les genres les plus rentables. Graphique utilisé : Boxplot des profits par genre.

Ce graphique montre que certains genres, comme l'action et la comédie, ont des distributions de profitabilité très variables, avec plusieurs films extrêmement rentables. À l'inverse, d'autres genres, comme les documentaires, affichent des profits plus modestes mais plus constants.

#### **Relation entre popularité et profitabilité**

Objectif : Évaluer l'influence de la popularité d'un film sur sa rentabilité. Graphique utilisé : Hexbin (carte de densité).

Les films avec une popularité élevée tendent à être plus rentables, mais des exceptions subsistent. Les blockbusters dominent la zone de profitabilité élevée, reflétant l'importance des campagnes marketing et des audiences globales.

#### **Distribution des profits selon la durée**

Objectif : Analyser si la durée d'un film influence ses profits. Graphique utilisé : Carte thermique.

Les films d'une durée autour de 100 minutes apparaissent souvent parmi les plus profitables. Cela pourrait refléter une préférence du public pour des durées modérées ou un équilibre optimal entre coûts de production et audience.

## **Impact de la langue parlée**

Objectif : Étudier l'effet de la langue des films sur leur rentabilité. Graphique utilisé : Boxplot des profits par langue.

Les films en anglais dominant en termes de profitabilité, ce qui reflète leur capacité à toucher un public mondial. Cependant, certaines langues régionales montrent des exceptions avec des valeurs extrêmes, représentant probablement des succès locaux.

## **Conclusion et perspectives**

Grâce à cette exploration, nous avons identifié plusieurs facteurs influençant la profitabilité des films :

1. Un budget plus élevé tend à augmenter les revenus mais pas toujours la rentabilité.
2. Les genres comme l'action et la comédie affichent une profitabilité variable mais prometteuse.
3. La popularité reste un indicateur clé pour prédire les succès financiers, mais elle n'est pas universellement garante de profits.
4. Une durée modérée de 100 minutes semble idéale pour maximiser les profits.
5. Les films en anglais bénéficient d'une audience plus large, ce qui booste leur profitabilité.

Ces résultats ouvrent la voie à des analyses approfondies pour comprendre comment optimiser la rentabilité des productions cinématographiques en fonction des tendances actuelles et des préférences des spectateurs.

## **3 Annexe**

**Q1**

**Q2**

**Q3**