Is There Disdain for Comedy?

Author : Noam Chabot
Date : January 2026

Introduction

In Western aesthetic thought, comedy has long occupied an ambivalent position. While tragedy was traditionally regarded as the privileged vehicle of truth, seriousness, and moral insight, comedy was often dismissed as secondary or frivolous. Since the Enlightenment, however, several authors have argued that comedy constitutes a philosophical paradigm in its own right, capable of reshaping conceptions of truth, morality, and critique (Kieran, 2013; Street, 2018). More broadly, the classical opposition between the tragic and the comic has progressively lost its rigidity, giving rise to hybrid narrative forms and shared expressive functions in modern cultural production (Mosse and Street, 2016).

Despite this theoretical rehabilitation, comedy remains a peculiar case in contemporary cinema. On the one hand, it is among the most popular genres with general audiences and occupies a central place in mass entertainment (Georgakopoulou, 2000). On the other hand, several studies suggest that comedy is critically undervalued, particularly in professional evaluations, which tend to privilege seriousness, artistic ambition, or formal innovation (Arell, 2012; Forshaw, 2022). Empirical research on film reception has further documented systematic divergences between audience preferences and expert judgments, with niche or artistically coded films often receiving greater critical attention and harsher evaluative standards than mainstream productions (Chen et al., 2021). In addition, expert reviews are not purely expressive but may also be shaped by reputational and strategic incentives, which can lead critics to distance themselves from popular tastes (Camara and Dupuis, 2014).

Within this literature, the question of whether comedy is evaluated less favorably than other genres has therefore been addressed primarily through the lens of reception, that is, by comparing audience ratings, critical scores, and commercial success. However, another dimension of symbolic valuation, institutional recognition through awards and festivals, has received comparatively less systematic attention. When awards are considered, existing work generally focuses on outcomes, comparing winners and nominees, rather than on the process through which films are selected for competition in the first place.

This focus on outcomes leaves an important gap in the literature. While differences between critics and audiences are well documented, and while the role of festivals and awards in shaping cultural legitimacy is widely acknowledged, little empirical research has examined whether comedy is already disadvantaged upstream, at the stage of institutional selection. In other words, it remains

unclear whether any bias against comedy operates through differential evaluation of competing films, or through a more structural filtering mechanism that limits comedy's access to the space of symbolic consecration altogether.

To what extent is comedy symbolically devalued in contemporary cinema, and through which mechanisms does this devaluation operate, including popular reception, critical evaluation, and institutional recognition?
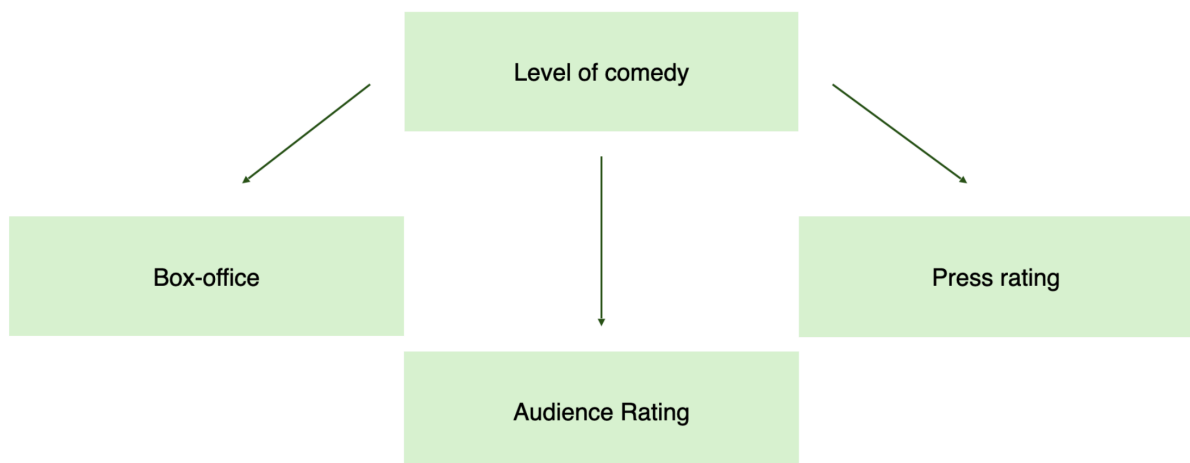
The present study addresses this gap by distinguishing explicitly between three stages of cinematic valuation: market performance and reception, award attribution, and institutional selection. Using large-scale datasets combining box-office data, audience and press ratings, genre-based measures of comedy, and large language model (LLM)-based classifications, we first examine how comedy relates to commercial success and evaluative judgments. We then analyze whether comedy influences award outcomes once films have been nominated. Finally, adopting a stratified bootstrapping approach, we test whether comedy is underrepresented among films selected for major French film institutions, the Festival de Cannes and the César Awards, relative to plausible pools of eligible films.

1. Comedies in box-office and reviews

1.1. Reviews and box-office with label classification

Are comedies rated less favorably by critics than by audiences?
To answer this question, we formulated three hypothesis:

Predictions:

• **P1** - The higher a movie's level of comedy, the greater its box-office success.

• **P2** - The higher a movie's level of comedy, the better its audience rating.

• **P3** - The higher a movie's level of comedy, the lower its rating from the press.

1.1.1 Data

We used two publicly available datasets from Kaggle for our analysis.
The first dataset (available here) includes N = 1,306,063 films from the TMDB database, containing classical information such as movie titles, release dates, and genres.
The second dataset (available here) contains N = 69,263 films with data on critical and audience ratings from Rotten Tomatoes, including the Tomatometer (an aggregated score based on professional critics' reviews, mostly from press, measuring the share of positive reviews out of all reviews submitted) and audience ratings. We treat the Tomatometer score as a measure of press or critical evaluation, since it is derived exclusively from professional critics' reviews, whose opinions tend to shape public perception and strongly influence audience behavior.

We merged the two datasets and retained only the films present in both sources.

We then removed rows with missing values in the Tomatometer score, the audience score, or the box-office revenue.

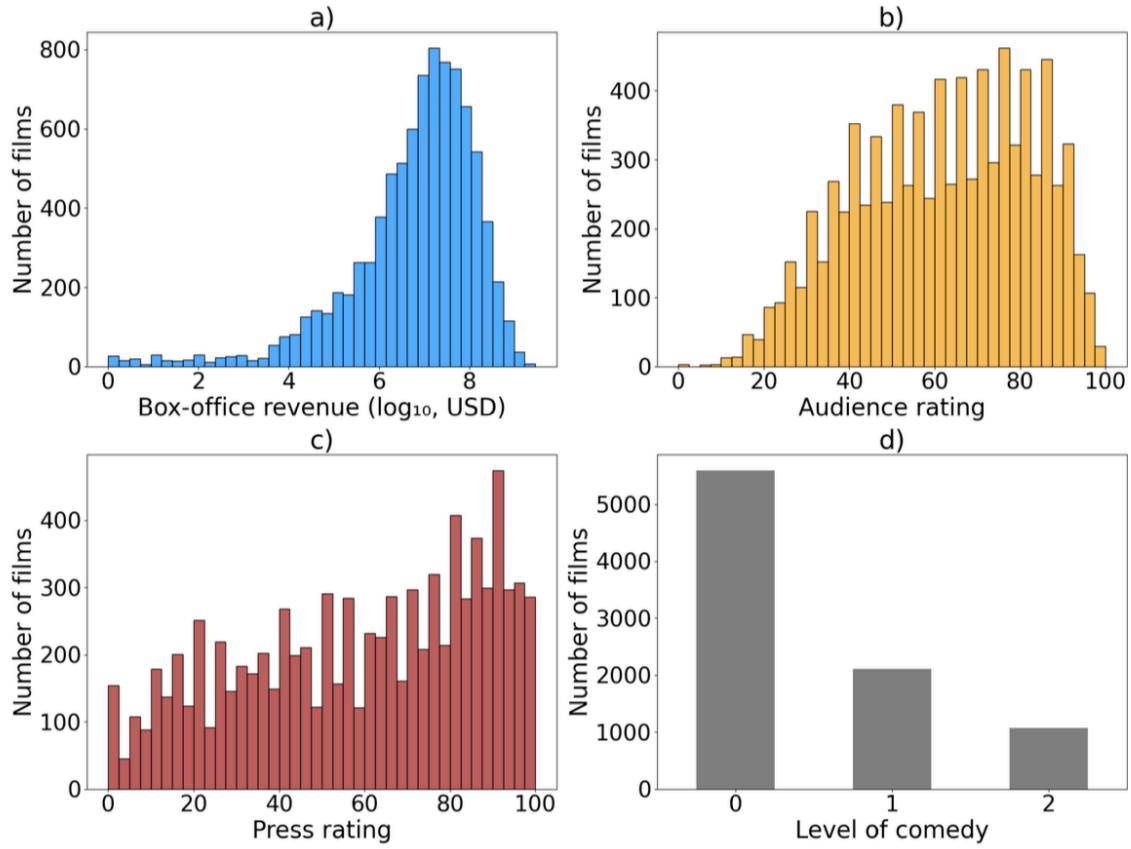After cleaning, the final dataset contained **N = 8,775** films, from 1913 to 2025.
Finally, we created a new variable measuring the "level of comedy" of each film, based on the TMDB genre tags, using the following algorithm:

- **0 (No Comedy)** – The Comedy tag is absent from the list of genres.

- **1 (Mixed Comedy)** – The Comedy tag is present alongside at least one of the following genres: Drama, Action, Romance, or Horror.

- **2 (Pure Comedy)** – The Comedy tag is present, and none of the four genres above appear.
These four genres were chosen because, in classical and neoclassical theatre traditions, they were considered incompatible with Comedy. For example, a comedy could not include elements of horror. The case of Romance is more nuanced: romantic comedies existed in classical theatre, but from a contemporary perspective, romantic comedy and pure comedy slightly differ in both narrative and humorous structure.

We used four variables of different types to test our predictions:
- Continuous: the **box-office revenue** in USD (**Figure 2a**).
- Continuous: the **audience rating** for each film (0-100, **Figure 2b**);
- Continuous: the **press rating** (0-100, **Figure 2c**);

- Ordinal: the **level of Comedy** (see the scale above, **Figure 2d**);

Figure 2 − Repartition according to the number of movies of a) the box-office revenue (log₁₀, USD), b) the audience rating (from 0 to 100), c) the press rating (from 0 to 100), d) the level of comedy given by labels (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy)

1.1.2. Methods

To test our three predictions, we compared box-office revenue, audience ratings, and press ratings as a function of the level of comedy (**Figure 3.d**, **3.e**, and **3.f**). We conducted three simple linear regressions, assuming equal variance and an equivalent gap between categories 0 and 1, and between 1 and 2. Because the distribution of revenues is highly skewed, with many low-revenue movies and only a few high-revenue ones, we applied a logarithmic (base 10) transformation to the revenue variable. The corresponding formula is:

$$Ri = b0 + b1Ci + \varepsilon i$$

Where :

R_i : the rating or revenue of movie i,

C_i : the comedy level of movie i,

ε_i : the error term.

It should be noted that the measure of comedy used in this study is deliberately coarse. By construction, the comedy variable takes only three ordinal values (no comedy, mixed comedy, pure comedy), and our analyses implicitly assume an equal distance between these categories. This assumption does not claim that comedic intensity is naturally discrete or linear, but rather constitutes a pragmatic operationalization that allows for consistent comparisons across large and heterogeneous datasets. Similar simplifications are common in large-scale cultural analyses, where interpretability and robustness often require a trade-off with measurement granularity. Importantly, this operational choice is applied symmetrically across all samples and comparisons, and therefore does not, in itself, bias results in favor of or against comedy.

1.1.3 Results

The regression shows that the level of comedy in a film increases significantly with its log box-office revenue (b = 0.239\$, $p < 0.001$, R^2 = 0.013\$). Conversely, the level of comedy decreases significantly with audience ratings (b = -2.440, $p < 0.001$, R^2 = 0.007) and press ratings (b = -4.690, $p < 0.001$, R^2 = 0.014).
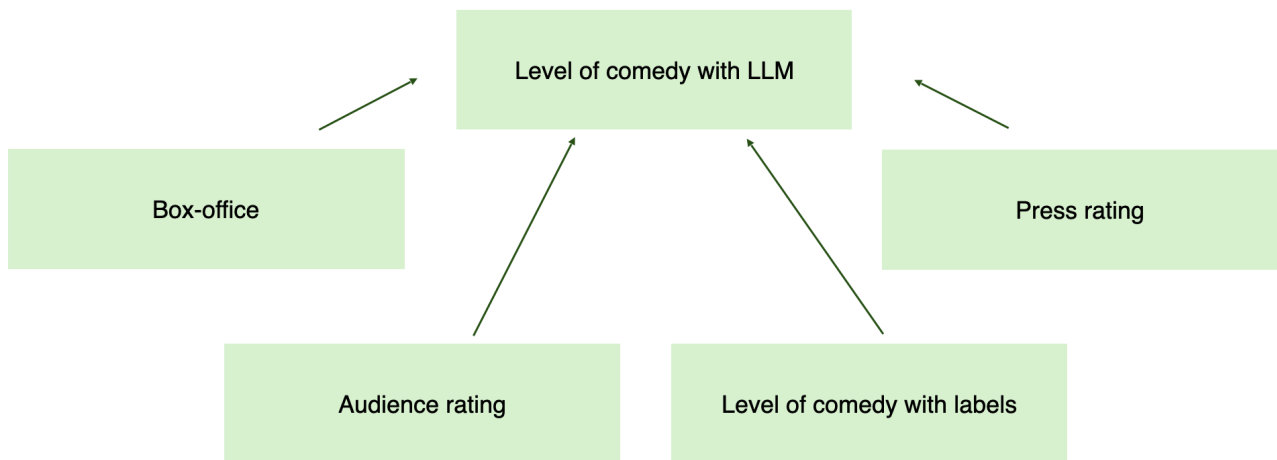
1.1.4. Conclusion

The results derived from the analysis of Rotten Tomatoes reviews and the TMDB database indicate that the degree of comedy in a film has a positive and statistically significant effect on box-office performance, while exerting a negative effect on critics' ratings. However, contrary to our initial hypothesis, this negative influence is also statistically significant for audience ratings. This finding suggests a divergence in audience profiles between box-office consumers and spectators who actively rate films, pointing to potentially distinct evaluative logics across these groups.

Nevertheless, a potential limitation of this study lies in its exclusive reliance on film genre labels. These labels are primarily assigned by distributors with marketing objectives in mind, favoring audience appeal over descriptive accuracy. For this reason, some studies have chosen to move beyond official genre labels by developing alternative methods, such as topic modeling using Latent Dirichlet Allocation or classification approaches based on movie posters (Chu et al., 2017; Chao et al., 2016). However, these methods ultimately depend on official labels for training purposes, which reintroduces the same underlying bias.

An alternative approach that avoids this issue, and which we develop in the present study, consists in using a large language model (LLM). Recent research shows that LLMs achieve performance comparable to more traditional models and can be particularly relevant for classification tasks (Wang et al., 2023). Their primary advantage in our context, which also constitutes their main limitation, is that they rely exclusively on latent knowledge. If genre information were incorporated alongside this latent knowledge, the model would be mechanically influenced by genre labels, thereby reproducing the original problem.

Accordingly, the approach adopted here is not intended to replace our previous analysis, but rather to complement it by providing an alternative method for estimating the degree of comedy in a film. We therefore replicate our analysis under the same hypotheses, substituting genre-based measures with estimates derived from an LLM.

1.2. Reviews and box-office with LLM classification

Predictions:

- **P1** - The higher a movie's level of comedy, the greater its box-office success.

- **P2** - The higher a movie's level of comedy, the better its audience rating.

- **P3** - The higher a movie's level of comedy, the lower its rating from the press.

- **P4 -** Films that receive higher comedy scores from a large language model also tend to exhibit higher comedy levels according to genre labels

1.2.1 Data

The only variable that differs from the previous study is the 'Comedy' variable, which is constructed using a different methodology. By definition, annotating data with a large language model (LLM) is computationally time-consuming, as each observation must be processed individually. Consequently, it is not feasible to annotate as many films as in our earlier analysis. We therefore set an annotation threshold of 500 films.

To maximize the likelihood that the LLM possesses latent knowledge of the selected films, we restricted our sample to the 500 most popular films according to TMDB's *Popularity* variable. Popularity on TMDB is a dynamic score computed on a daily basis by aggregating user interactions, including page views, votes, and additions to favorites or watchlists. The algorithm also incorporates a recency factor, such that popularity scores naturally decay over time in order to emphasize current trends. Based on this mechanism, we assume that the LLM is more likely to have internalized information about these highly popular films than about less visible ones.

For the classification task, we made a trade-off between the LLM's latent knowledge capacity and its number of parameters, as larger models cannot be run efficiently on all machines or on commonly used cloud services such as Google Colab (free tier). After reviewing several benchmark results, we ultimately selected **Qwen2.5-7B** as a compromise between performance and computational feasibility.

We initially designed a prompt and tested it on a small subset of films. A recurring issue was that Qwen tended to assign excessively high comedy scores (often 10/10) to films that are better characterized as mixed-genre comedies. Moreover, the model appeared to infer or guess rather than rely strictly on its latent knowledge, which ran counter to our methodological objectives. We therefore revised the prompt to address these issues, incorporating explicit constraints designed to minimize hallucination and speculative reasoning. The final prompt is presented below.

You are an expert film critic and genre analyst. Your task is to assign a "Comedy Score" (from 1 to 10) to the following movie.

MOVIE: "{movie_title}"

CORE PRINCIPLE:
Assess the **intent** of the movie to provoke laughter. Do not judge quality, but the density and purity of humor.

RATING SCALE (Strictly follow these nuances):
- **10 (Undiluted Comedy)**: The ONLY goal is laughter. The film **cannot** be classified as another genre (Action, Romance, Drama). The plot is merely a vessel for gags, farce, or absurdity. (e.g., Airplane!, Monty Python).
- **8-9 (Genre-Hybrid Comedy)**: Extremely funny, BUT the movie also functions as another genre. It is an Action-Comedy, a Romantic-Comedy, or a Satire. The stakes feel somewhat real despite the jokes. (e.g., Deadpool, Barbie, There's Something About Mary).
- **5-7 (Balanced Dramedy / Action)**: Humor is a major ingredient (50%), but shares the spotlight equally with serious plot points, tension, or character drama. There is a solid narrative backbone

- **3-4 (Serious with Comic Relief)**: The movie is a Thriller, Adventure, or Drama first. Humor is used only to break tension, but the main story is serious.
- **1-2 (Serious / Dark / Pure Drama)**: No intent to be funny. The atmosphere is grave, tragic, or strictly factual. Even if the dialogue is intelligent ("witty"), if the goal isn't laughter, it belongs here.

IMPORTANT RULES:
1. **The "Hybrid" Ceiling:** If a movie relies on Action sequences (fights) or a Romance arc to move the story forward, the maximum score is **9**. A 10 is reserved for pure structural chaos/comedy.
2. **Wit vs. Comedy:** Do not confuse clever dialogue with comedy. A serious drama with sharp dialogue is a 1 or 2 or 3.
3. **Unknown Movies:** If you do not know the movie enough, strictly output 0. Do not try to guess.

RARITY & EXCEPTION RULE:
A score of 10 should apply to at most ~1-2% of all narrative feature films ever made.
Only assign 10 or 9 if removing comedy would destroy the film entirely.
If the film still functions as Action, Romance, Drama, or Satire without jokes, the maximum score is 7.
When in doubt between two numbers, always choose the lowest.

RESPONSE FORMAT (Strictly follow this line-by-line structure):
Line 1: Knowledge Level: [High/Medium/Low/None]
Line 2: Justification: [Your 2-3 sentences analysis]
Line 3: ### Note : [Integer only]

As shown by the prompt design, we explicitly distinguish structural comedy, in which the primary narrative function is to generate laughter, from comic relief, which serves merely to alleviate tension within an otherwise serious or dramatic storyline. The prompt operationalizes this distinction through a finely graded rating scale and a set of restrictive rules that prioritize authorial intent over perceived quality, while also discouraging speculative judgments when the model lacks sufficient knowledge. In particular, the inclusion of explicit ceiling effects, rarity constraints, and a strict "unknown movie" rule was intended to force conservative scoring behavior and minimize overestimation.
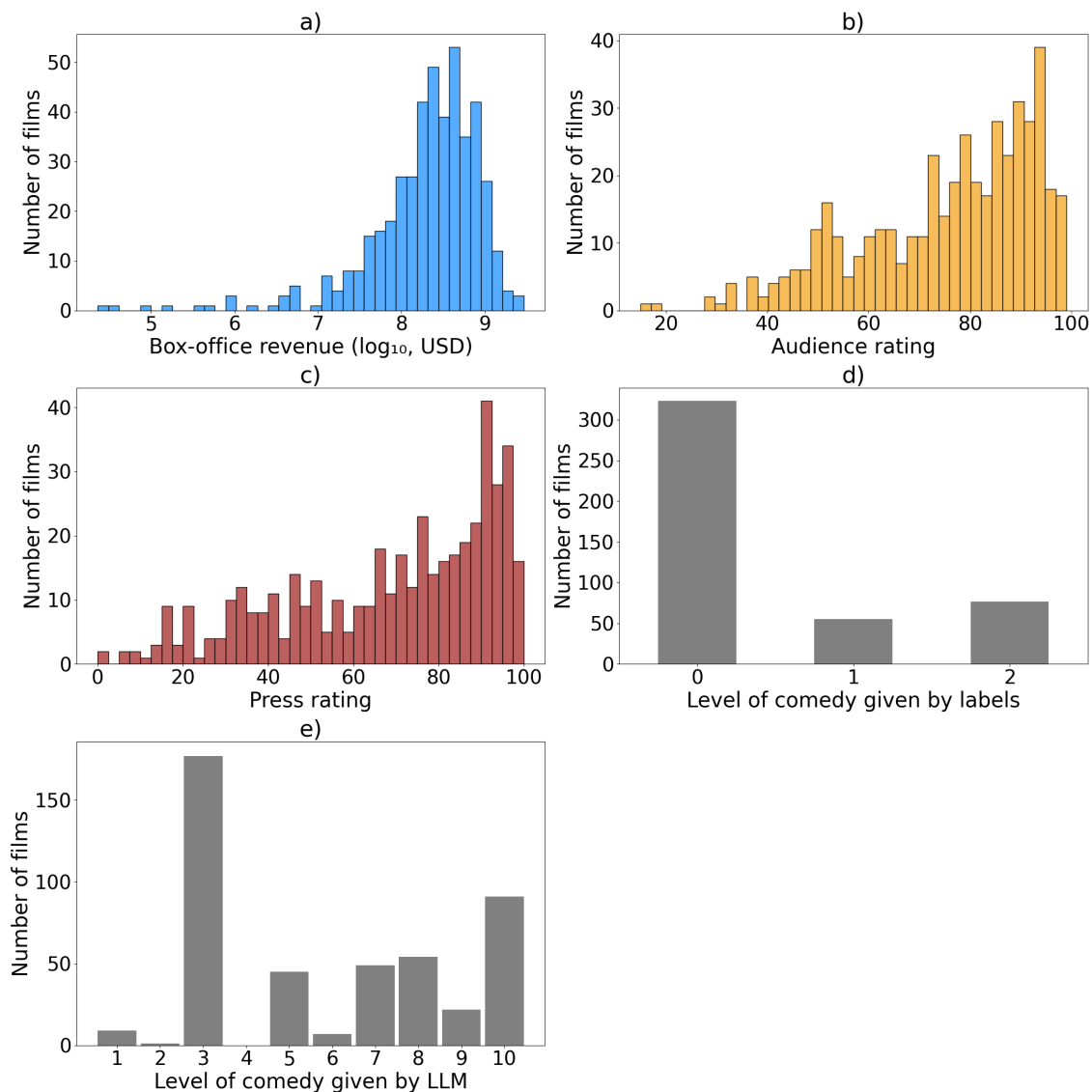
Using this prompt, we inferred comedy scores for the selected sample of 500 films and collected, for each observation, the numerical score, a short textual justification, and an explicit self-reported level of certainty regarding the model's knowledge of the film.

To further ensure that the model did not rely on conjecture or hallucinated information, we applied a series of post-processing filters. Specifically, we removed all observations with a score of 0, all cases in which the model's reported knowledge level was not *High*, and all responses containing lexical markers of uncertainty such as *"appears"* or *"appeared"*. This filtering process resulted in the exclusion of 60 observations, yielding a final dataset of N = 440 films.

The post-processing filters applied to the LLM outputs reflect a deliberate trade-off between validity and representativeness. By retaining only films for which the model explicitly reports a

high level of knowledge and by excluding responses containing markers of uncertainty, we aimed to minimize hallucination and speculative inference, which are well-documented limitations of large language models. However, this conservative strategy may also introduce selection effects, as it likely favors well-known or canonical films and underrepresents less visible productions. As a result, the final sample may not fully capture the diversity of genre hybrids or marginal comedies. This limitation should be taken into account when interpreting the LLM-based results, which prioritize reliability of classification over exhaustiveness.

Qwen produced relatively heterogeneous scores: while there is no strong overall overrepresentation of numerical values, the distribution remains asymmetric for certain digits, most notably an overrepresentation of the value 3 and the complete absence of the value 4 (see **Figure 1.e**).
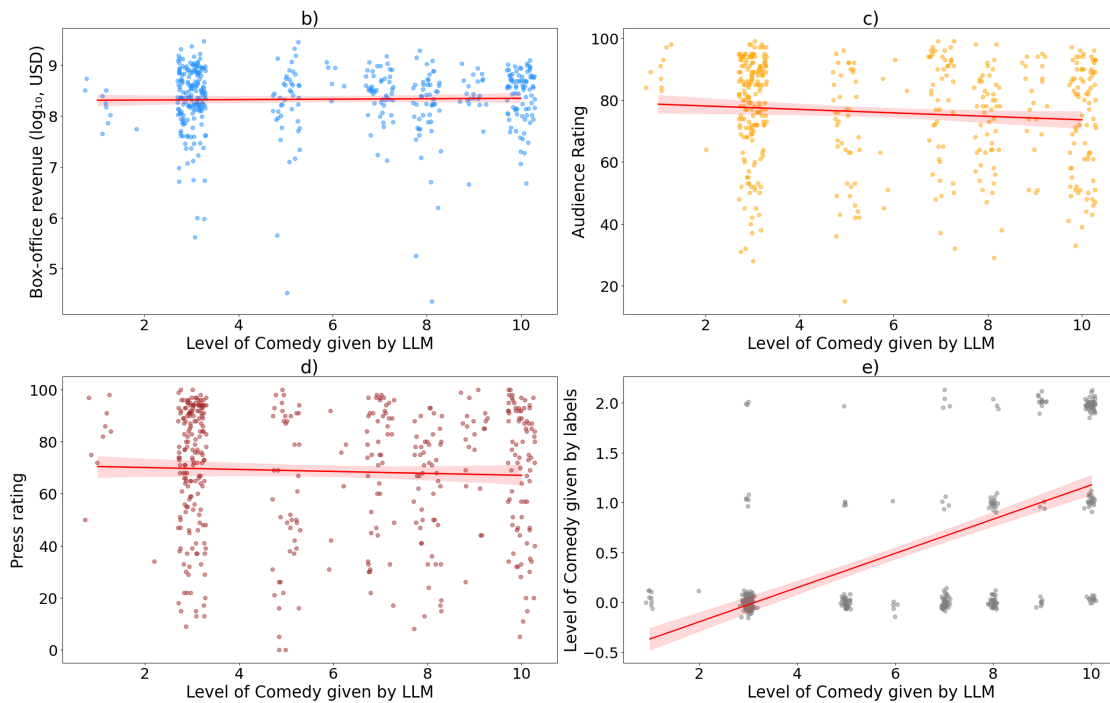

Figure 1 - Repartition according to the number of movies of a) the box-office revenue ($\log_{10}$, USD), b) the audience rating (from 0 to 100), c) the press rating (from 0 to 100), d) the level of comedy given by labels (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy) and e) the level of comedy given by the Qwen2.5-7B model

1.2.2 Method

We used exactly the same method as the previous tests.

1.2.3 Results

The regression shows that the level of comedy in a film increases slightly with its log box-office revenue (b = 0.004, R^2 = 0.000, see **Figure 2.b**), but the result is not significant (p = 0.698). Moreover, the level of comedy decreases with its press rating (b = -0.378, R^2 = 0.002, , see **Figure 2.d**), but the result is not significant neither (p = 0.359). However, the level of comedy decreases with its audience rating and the result is significant (b = -0.560, p < 0.05, R^2 = 0.009, see **Figure 2.c**), and also significantly with the level of comedy given by the label (b = 0.172, p < 0.001, R^2 = 0.415, see **Figure 2.e**)

Figure 2 – Repartition of b) the box-office revenue ($\log_{10}$, USD), c) the audience rating (from 0 to 100), and d) the press rating (from 0 to 100) and e) all depending on the level of comedy given by LLM (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy)

1.2.3 Conclusion

Overall, the regression indicates that the level of comedy is not significantly associated with box-office revenue or press ratings, whereas it is significantly related to audience ratings and to the comedy level derived from genre labels, with comedy being lower for films with higher audience ratings and strongly aligned with the label-based measure.

The fact that LLM-based variables are overall less statistically significant can, in our view, be explained by three main factors.

First, a purely mechanical explanation cannot be ruled out. Given the substantially smaller dataset, statistical tests naturally have lower power, which may result in non-significant estimates. Importantly, although the coefficients fail to reach significance, their directions (positive or negative) are consistent with those observed in our previous label-based study, suggesting that the underlying relationships may still be present but insufficiently supported by the available data.

Second, it is possible that the LLM provides a more faithful representation of the intrinsic comedic content than genre-based labels. Under this interpretation, our previously validated hypothesis—that comedy positively influences box office performance and negatively influences press ratings— would need to be reconsidered, as these effects are no longer statistically supported. This would imply that the effects identified in the label-based analysis may partly reflect extrinsic factors associated with genre categorization rather than comedy as an inherent cinematic property.

Third, and this is the hypothesis we favor, the LLM may be less reliable than genre-based classification for this task. The weaknesses observed during the training phases do not allow us to rule out additional errors occurring at inference time. In particular, the well-documented tendency of LLMs to generate plausible but incorrect outputs, even when strong warnings are included in the prompt, remains a critical concern. Moreover, zero-shot LLM-based studies typically rely on very large models (Wang et al., 2023), whereas the models used here may not reach the same level of robustness.

In addition, it should be noted that the genre label displayed to audiences may itself influence viewer behavior. Genre information can shape expectations and guide film choice, meaning that labels may exert an indirect but substantial effect on box office outcomes. This mechanism could explain why the results observed with genre-based classifications are not recovered when relying solely on LLM-inferred attributes.

Finally, a potential source of bias in this study lies in the selection of the 500 most popular films. If this subset overrepresents certain genres or systematically includes films with a higher volume of reviews, the resulting estimates may be distorted. Such a selection bias could attenuate or amplify observed effects, particularly if popularity correlates with both genre and critical attention. Plus, we can visually see that the distribution of the 500 most popular films is quiet different than the original one (see **Figure 1 a**, **b**, **c** and **d**). Although we do not believe this bias directly drives the present results, it remains an important limitation that should be acknowledged.

Transition

After examining box-office performance and critical reception, we now turn to the quantification of the presumed underrepresentation of comedy in French film festivals, focusing on the two major ones: the Cannes Film Festival and the César Awards.

2. Comedies in French film festivals

2.1 Comedies in Césars Awards

Is there a bias against comedy in French film festivals?

To answer this question, we formulated two hypothesis:

Predictions:

- **P1** - The winners of the French Academy Award for Best Picture tend to have a lower level of comedy than the nominated films.

- **P2** - The winners of the French Academy Award for Best Director tend to have a lower level of comedy than the nominated films.

2.1.1 Data

We constructed two datasets based on films nominated for the César Awards.

The first dataset includes all films nominated for the Best Picture category from 1976 to 2025 (N = 271), along with additional information such as the film's director, nomination status (nominated or winner), and the Wikipedia pages of both the film and its director. The second dataset contains the same information for the Best Director category (N = 265), covering the same time period (1976–2025).

Both datasets were compiled by extracting data from the HTML source code of the corresponding Wikipedia pages (César du meilleur film and César de la meilleure réalisation), accessed on October 3, 2025.

Next, for each film, we added its synopsis and genre tags by scraping data from its individual Wikipedia page via the Wikipedia API on November 6, 2025.

Finally, we concatenated the two datasets so that each film appearing in either category (or both) was included.

Each film therefore contained the two variables 'Best Picture' and 'Best Director', each with three possible values: Winner, Nominated, Not Nominated (if the film was nominated only in the other category).

However, we decided to remove the Not Nominated cases from our tests, as they contributed little to assessing the validity of our two hypotheses.

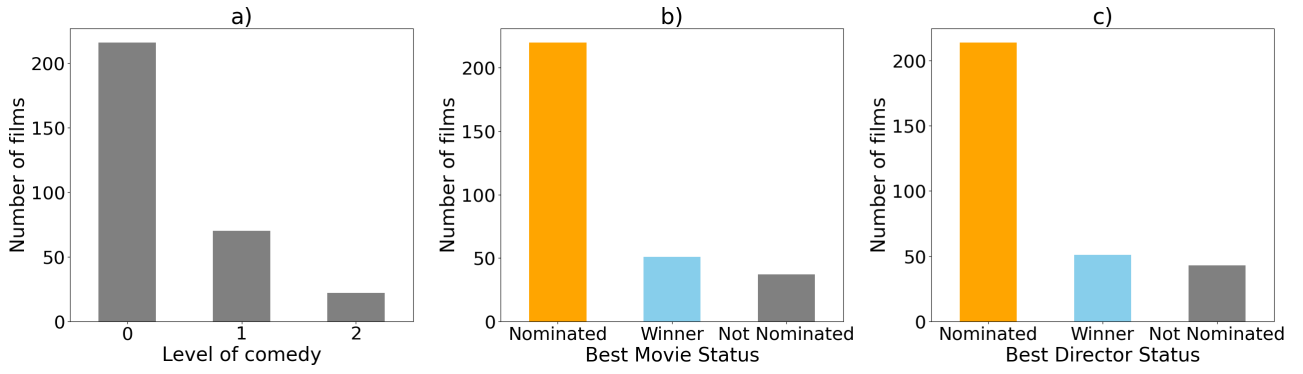The final merged dataset included **N = 308 films**.

Finally, we created a new variable measuring the "level of comedy" of each film, based on the Wikipedia genre tags, using the following algorithm:

- **0 (No Comedy)** – The Comedy tag is absent from the list of genres.

- **1 (Mixed Comedy)** – The Comedy tag is present alongside at least one of the following genres: Drama, Action, Romance, or Horror.

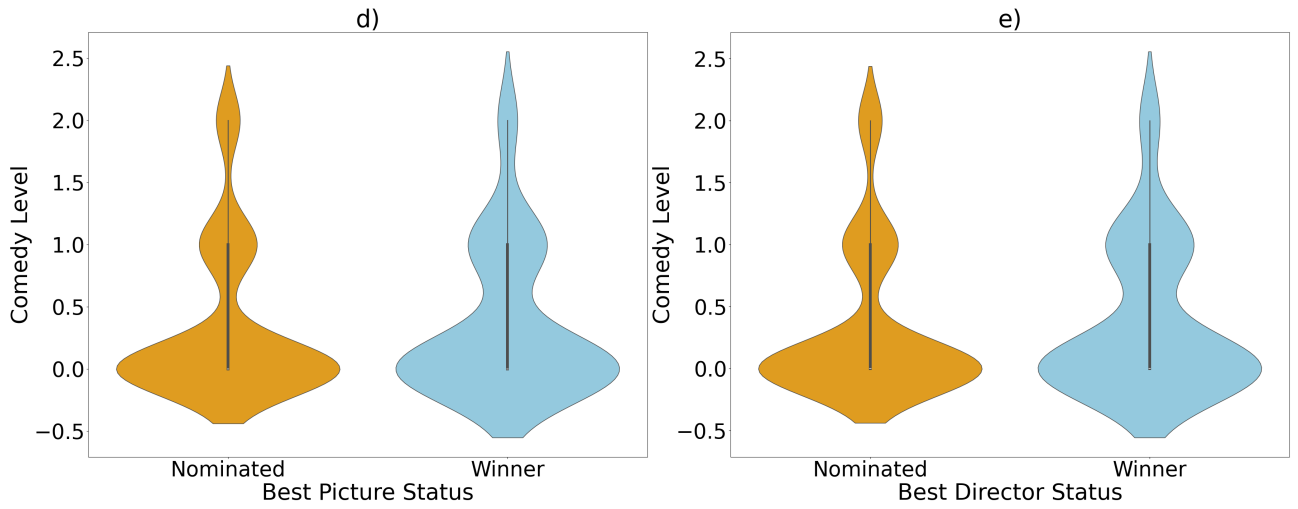- **2 (Pure Comedy)** – The Comedy tag is present, and none of the four genres above appear.

These four genres were chosen because, in classical and neoclassical theatre traditions, they were considered incompatible with Comedy. For example, a comedy could not include elements of horror. The case of Romance is more nuanced: romantic comedies existed in classical theatre, but from a contemporary perspective, romantic comedy and pure comedy slightly differ in both narrative and humorous structure.

We used three main variables of different types to test our predictions:

- Ordinal: the **level of Comedy** (see the scale above, **Figure 2a**);

- Binary: the film's status in the **Best Picture** category (Winner or Nominated, **Figure 2b**)

- Binary: the film's status in the **Best Director** category (Winner or Nominated, **Figure 2c**)

Figure 2 - Repartition according to a) the level of comedy (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy), b) the movie status at the César Awards for the best Picture and c) the movie status at the César Awards for the best Director

Figure 3 - Repartition of the level of comedy (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy) according to d) the movie status at the César Awards for the best Picture and e) the movie status at the César Awards for the best Director

2.1.2 Methods

To test our two predictions, we compared the level of comedy in movies depending on whether they won the Best Picture award or not, and whether they won the Best Director award or not (**Figure 3.d** and **Figure 3.e**). We conducted two independent **t-tests**, assuming equal variance and an equivalent gap between categories 0 and 1, and between 1 and 2. The corresponding formula is:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Where :

μ_1 and μ_2 : mean comedy levels of each group

s_p^2 : pooled variance

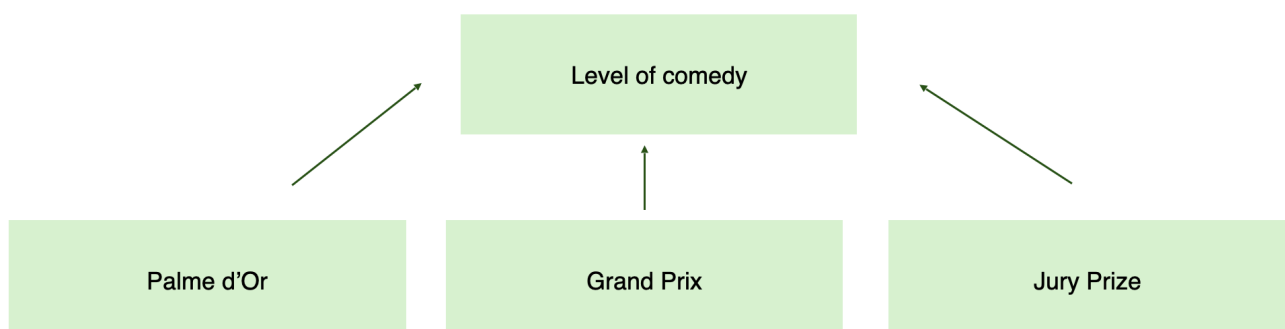n_1 and n_2 : sample sizes of the two groups

2.1.3 Results

The mean of the comedy level of movies among the winners of the best Picture is slightly higher than for the nominated (μ_1= 0.367, μ_2= 0.346, see **Figure 2.d**), but this difference is not significant (t = -0.215, p = 0.831). Plus, the mean of the comedy level of movies among the winners of the best Picture is slightly lower than for the nominated (μ_1= 0.388, μ_2 = 0.341, see **Figure 2.e**), but this difference is not significant neither (t = -0.478, p = 0.634)

2.1.4 Conclusion

These results indicate that there is no significant difference in the level of comedy between winning and nominated films in either case, suggesting that comedy is neither favored nor penalized in award outcomes within this sample. In other words, the presence of comedic elements does not appear to play a decisive role in distinguishing winners from nominees.

However, the Cannes Film Festival is widely regarded as more elitist in its artistic positioning, which makes it particularly interesting to investigate whether a differential effect of comedy could be observed in that context, in contrast to the absence of such an effect in the present results.

2.2 Comedies in the Festival de Cannes

Predictions:

- **P1** - The winners of the Palme d'Or tend to have a lower level of comedy than the official Selection.

- **P2** - The winners of the Grand Prix tend to have a lower level of comedy than the nominated films.

- **P3** - The winners of the Jury Prize tend to have a lower level of comedy than the nominated films.

2.2.1 Data

We constructed the dataset using the Wikipedia API.

We first used the Wikipedia MediaWiki API to check, for each year since the creation of the Festival de Cannes (i.e., from 1946 to 2025), whether a page with the exact title "YYYY Cannes Film Festival" exists, automatically following redirects.
When such a page is found, the main metadata (year, canonical title, page ID, and full URL) are collected into a structured list. The only two years for which no Wikipedia page exists are 1948 and 1950, which is consistent with the fact that the festival was cancelled in those years.

We then scraped, for each existing page, the full list of films competing in the Official Competition. The Official Competition should be distinguished from the Official Selection, which includes all films chosen by the festival to be presented under its official label and encompasses several distinct sections, such as the Official Competition, Un Certain Regard, and Out of Competition screenings. While all these films are part of the Official Selection, only those in the Official Competition are eligible for the Palme d'Or. The Official Competition is therefore a subset of the Official Selection, restricted to films competing for the main awards, including the Palme d'Or. We chose to focus on this category because it is the most prestigious and the one most prominently highlighted in film promotion and media coverage.

We then enriched each film with a synopsis through additional scraping, as well as with metadata such as the director, the genre as provided by Wikipedia, and the film's country of origin using the API.

We then created three additional datasets by scraping the respective Wikipedia pages that list all winners in the Palme d'Or, Grand Prix, and Jury Prize categories. These three categories were selected because they are considered major awards and are attributed to films rather than to individuals, unlike, for example, the Best Director Award. The Palme d'Or represents the festival's highest distinction and is awarded to the best film in the Official Competition. The Grand Prix is the second most important award, while the Jury Prize is a collective distinction recognizing a film for its originality or artistic singularity. We did not include the Best Director Award, as our analysis focuses on films themselves and on the overall cinematic work presented to audiences, rather than

on the individual director. All the manipulation and scraping on Wikipedia had been done on 6th January 2026.

Finally, we concatenated these three sub-datasets with the main dataset using the film's Wikipedia URL as the key, as this link is unique. We removed any rows corresponding to the years 1948 and 1950. We then addressed an important specificity of the Festival de Cannes: the major awards have not been awarded every year since the festival's inception. For instance, the Grand Prix in its current form, that is, as the second most important award after the Palme d'Or, was introduced in 1967, and the Jury Prize has not been awarded in some years. Consequently, we created three variables, Palme d'Or, Grand Prix, and Jury Prize, which take the value *Official Competition* if the film competed in that category, the name of the prize if the film won it, and *NaN* if the prize was not awarded in that year.

We therefore obtain a final dataset that includes all films selected for the Official Competition and indicates, for each major category, whether the film won the award or not, with the constraint that a film cannot win more than one major prize, as the competition rules impose exclusivity among major awards. The dataset will bd published on *Kaggle* soon, as no existing dataset provided a comprehensive and well-structured record of all films selected for the Official Competition, and even the few datasets listing Palme d'Or winners were poorly formatted.

For the purposes of our specific project, we decided to exclude films selected in 1968 and 2020, as no awards were granted in those years due respectively to the events of May 1968 and the COVID-19 pandemic. We also removed films for which genre information was missing, as this variable is central to the comparisons conducted in our analysis.

Finally, we created a new variable measuring the "level of comedy" of each film, based on the Wikipedia genre tags, using the following algorithm:

- **0 (No Comedy)** – The Comedy tag is absent from the list of genres.

- **1 (Mixed Comedy)** – The Comedy tag is present alongside at least one of the following genres: Drama, Action, Romance, or Horror.

- **2 (Pure Comedy)** – The Comedy tag is present, and none of the four genres above appear.
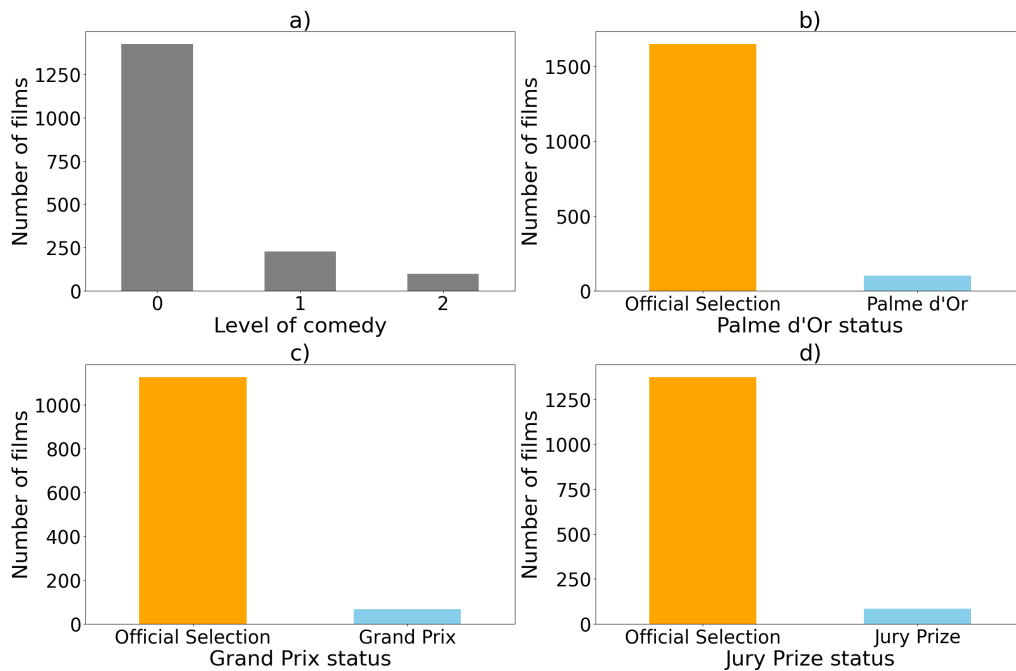
These four genres were chosen because, in classical and neoclassical theatre traditions, they were considered incompatible with Comedy. For example, a comedy could not include elements of horror. The case of Romance is more nuanced: romantic comedies existed in classical theatre, but from a contemporary perspective, romantic comedy and pure comedy slightly differ in both narrative and humorous structure.

The final dataset included **N = 1,802 films**.

We used three main variables of different types to test our predictions:

- Ordinal: the **level of Comedy** (see the scale above, **Figure 2a**);

- Binary: the film's status in the **Best Picture** category (Winner or Nominated, **Figure 2b**)

- Binary: the film's status in the **Best Director** category (Winner or Nominated, **Figure 2c**)

Figure 2 - Repartition of films according to a) the level of comedy (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy) and the film status at the Festival de Cannes for b) the Palme d'Or, c) the Grand Prix and d) the Jury Prize

2.2.2. Methods

To test our three predictions, like the César Award, we compared the level of comedy in movies depending on whether they won the Palme d'Or, Grand Prix or Jury Prize, or not (**Figure 3.d, 3.e and 3.f**). We conducted two independent **t-tests**, assuming equal variance and an equivalent gap between categories 0 and 1, and between 1 and 2. The corresponding formula is:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$
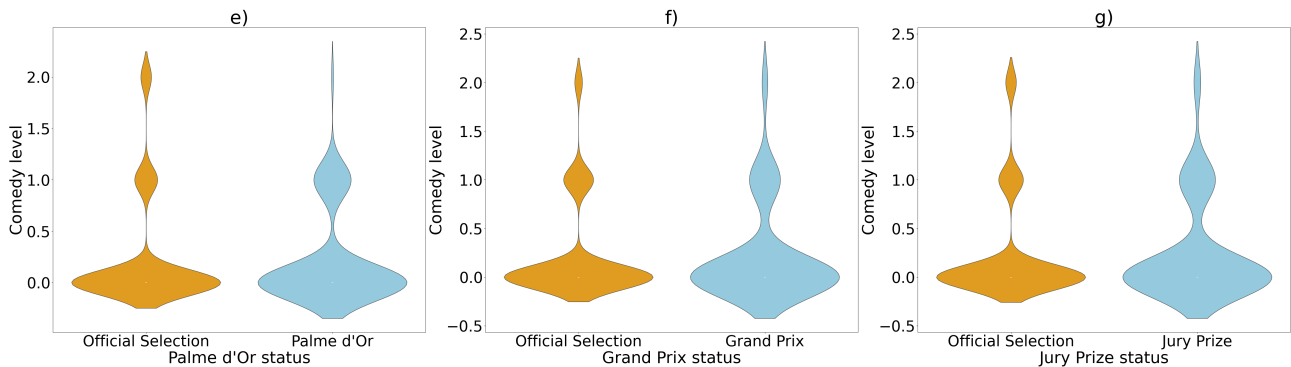
Where :

μ_1 and μ_2 : mean comedy levels of each group

s_p^2 : pooled variance

n_1 and n_2 : sample sizes of the two groups

2.2.3 Results

The mean of the comedy level of films among the winners of the best Picture is slightly higher than for the nominated ($\mu_1 = 0.218$, $\mu_2 = 0.242$, see **Figure 3.e**), but this difference is not significant (t = 0.523, p = 0.602). Plus, the mean of the comedy level of movies among the winners of the Grand Prix is also slightly lower than for the nominated ($\mu_1 = 0.227$, $\mu_2 = 0.242$, see **Figure 3.f**), but this difference is not significant neither (t = 0.233, p = 0.816). To finish, the mean of the comedy level of films among the winners of the Grand Prix is slightly higher than for the nominated ($\mu_1 = 0.259$, $\mu_2 = 0.249$, see **Figure 3.g**), but this difference is not significant neither (t = -0.175, p = 0.862)

Figure 3 - Repartition of the level of comedy (with 0: No Comedy, 1: Comedy mixed with another genre, 2 : Pure comedy) according to e) the Palme d'Or status, f) the Grand Prix status and g) the Jury Prize status

2.2.4 Conclusion

Across all comparisons, there is no significant difference in the level of comedy between winning and nominated films, whether for Best Picture or the Grand Prix. Overall, comedy does not appear to systematically is associated with award outcomes, as winners and nominees exhibit comparable levels of comedic content.

These results are interesting, but they do not allow us to answer our initial question regarding the potential underrepresentation of comedy in film festivals. To address this issue, it would be necessary to consider all films eligible for Cannes and the César Awards and to identify a bias that would result in comedy being underrepresented. We therefore adopted a similar approach using a bootstrapping technique and formulated two new hypotheses that more directly address our research question.

2.3 Comedies, Festivals and bootstrapping

Predictions:

• **P1** - Films nominated for the Cannes Film Festival exhibit a lower level of comedy than the full set of eligible films

• **P2** - Films nominated for the César Awards exhibit a lower level of comedy than the full set of eligible films

2.3.1 Data

With regard to film festivals, we reused the same data previously collected for the Cannes Film Festival and the César Awards. However, we reintegrated films nominated in 2020 and 1968 for the Cannes Film Festival, since our analysis focuses exclusively on nominees rather than winners; these years are therefore relevant to our hypothesis.

As a comparison baseline, we chose to use the TMDB dataset, which contains $N = 1,306,063$ films in its raw, unmanipulated form. The main challenge lies in identifying all films eligible for both festivals, as their eligibility criteria are not exactly the same. We operationalize eligibility as a plausibility set approximating institutional constraints rather than an exact legal eligibility list.

First, we processed the TMDB dataset to match the criteria common to both festivals. We removed films labeled with the genre "Documentary" as well as those marked "yes" in the *Adult* column, as these types of films are not typically nominated at festivals. We also restricted the dataset to films with a runtime exceeding 60 minutes, an explicit requirement of both festivals for a film to be considered a feature-length production.

Moreover, films selected at Cannes or nominated for the César Awards typically have a minimum level of industrial existence and receive a theatrical release (at least in France after selection). As a result, they are generally covered by the specialized press. We therefore assume that a film selected at Cannes has received at least 10 votes on TMDB. Films with fewer than 10 votes are often titles that were never truly released, or that correspond to amateur or local productions, television films, or database errors. In all cases, such entries do not match the profile of films selected at Cannes.

We therefore restrict the comparison set to films that have received at least 10 votes on TMDB, in order to exclude entries lacking minimal distribution or public visibility. This criterion allows us to better approximate the pool of films eligible for Cannes and the César Awards, without introducing a strong bias toward widely successful films. Finally, we further restricted the dataset to films that were actually released theatrically.

Cannes-specific criteria.
Feature-length films submitted for selection at the Cannes Film Festival must in particular meet the following requirements: they must have been produced within the twelve months preceding the festival; they must not have been commercially exploited outside their country of origin; they must not have been presented at another international film festival; and any feature-length film invited to compete must receive a commercial theatrical release in France, in compliance with French regulations, particularly those governing release windows. Unfortunately, it is difficult to verify that all of these criteria are fully satisfied using our available data.

As a pragmatic solution, we focused on the production country. We retained only the first listed country in cases of co-production. This choice mirrors our Cannes dataset, which also sometimes includes multiple countries, and simplifies alignment across datasets. We assume that co-producing countries are listed in the same order in both datasets, with the first country corresponding to the

principal producer. After applying these restrictions, the resulting dataset contains $N = 59,913$ films (see **Figure 1.a**, **1.b**, **2.e** and **2.f**).
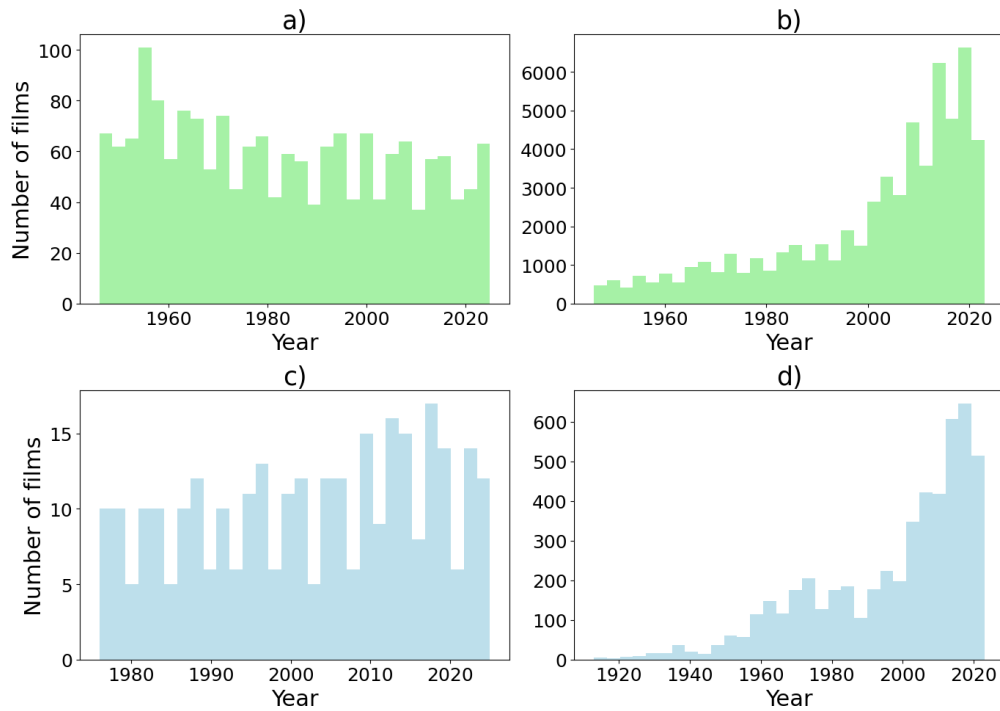
César-specific criteria.
A feature-length film is eligible for the César Award for Best Film if it meets all of the following conditions: it must be a feature-length film of French initiative (i.e., majority or relative-majority French production, regardless of shooting language); it must have received CNC production approval; it must hold a theatrical exhibition visa issued by the French Ministry of Culture (excluding temporary or exceptional visas); it must have had its first commercial theatrical release in at least one commercial cinema in the Paris region between January 1 and December 31 of the year preceding the ceremony; and it must have been commercially exhibited for at least one week (seven consecutive days) in that cinema.
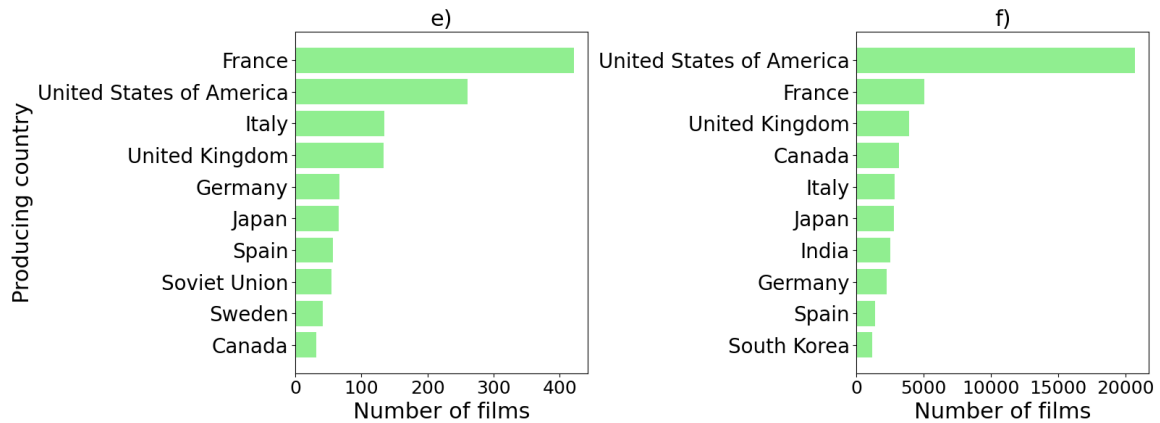
Once again, we do not have access to all of this information within the TMDB dataset. However, unlike Cannes, where films may be international, we are able to account for French production in the case of the César Awards. We therefore restricted the dataset to films whose country of production is France. Because films must be primarily produced in France, we again retained the first listed country in cases of co-production, treating it as the principal producer. After applying these constraints, the final dataset contains $N = 5,187$ films (see **Figure 1.a**, and **1.b**).

These operational choices reflect pragmatic approximations necessitated by data availability constraints in TMDB. While several official eligibility criteria for Cannes and the César Awards cannot be directly verified, the applied filters are intended to approximate as closely as possible the institutional selection pools of both festivals. Importantly, these restrictions are symmetrical with respect to genre and do not specifically disadvantage comedy, thereby limiting the risk of introducing systematic bias. As such, the resulting datasets should be interpreted as conservative but plausible representations of the films realistically eligible for selection at Cannes and the César Awards.
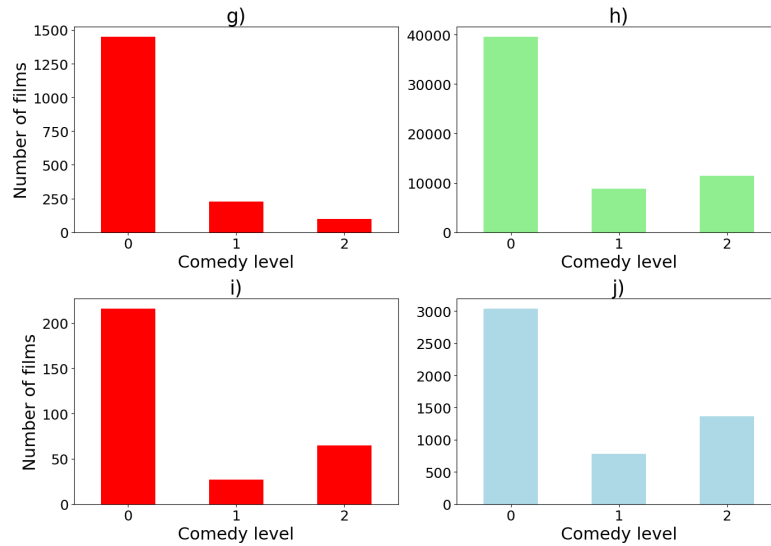
To measure the level of comedy, we chose to use the mean value of the comedy metric ranging from 0 to 2 that was already implemented (see **Figure 3.g**, **3.h**, **3.i, 3j**). We therefore assume that a pure comedy is "twice as comedic" as a mixed-genre comedy. This assumption was adopted pragmatically in order to simplify the calculations and to rely on a single statistical test, and it is justified in the same manner as in our previous analyses of the Cannes Film Festival and the César Awards. In addition, this approach is considerably easier to implement than an LLM-based method when dealing with a large number of films, as is the case here.

Figure 1 - Distribution of films by release year for a) César Award nominees, b) the corresponding TMDB comparison sample, c) Cannes Film Festival selections, and d) the corresponding TMDB comparison sample.

Figure 2 - Distribution of the ten most frequent producing countries for e) César Award nominees and f) the corresponding TMDB comparison sample

Figure 3 - Distribution of films by comedy level (0 = no comedy, 1 = mixed-genre comedy, 2 = pure comedy) for g) César Award nominees, h) the corresponding TMDB comparison sample, i) Cannes Film Festival selections, and j) the corresponding TMDB comparison sample.

2.3.2 Method

To test whether the official selection of the Cannes Film Festival exhibits a level of comedy different from that observed in the overall film population, we compare the observed statistic to a null distribution obtained via stratified bootstrapping from a reference dataset (TMDB). Strata are defined by year and main country of production. Each film is associated with a year y and a country c, and the Cannes selection is summarized by the number $n_{y,c}$ of films in each (y,c) cell. At each bootstrap iteration b, we construct a pseudo Cannes sample by randomly drawing, for each stratum (y,c), exactly $n_{y,c}$ films from TMDB belonging to the same stratum. Because films selected at Cannes are often released after their production year, we allow a temporal tolerance by including films released in year y and $y+1$. For example, if in 2005 the Cannes selection includes 10 French films and 3 Brazilian films, we randomly draw 10 French films released in 2005 or 2006 and 3 Brazilian films released in 2005 or 2006 from TMDB. Sampling is performed without replacement when the TMDB pool is sufficiently large, and with replacement otherwise, in order to preserve stratum sizes.

To handle rare strata that are present in Cannes but absent from TMDB, we apply a hierarchical fallback strategy. First, if the exact (y,c) stratum is empty in TMDB, we expand the temporal window to $[y-2, y+3]$ while keeping the country constraint. If this expanded pool is still empty, we relax the country constraint and draw from TMDB films released in year y only. After aggregating all draws across strata, we compute the statistic of interest T_b on the pseudo sample. Repeating this procedure B times yields an empirical distribution of T_b representing the null hypothesis that the Cannes selection is a conditionally random sample given the year by country structure. Statistical significance is assessed by comparing the observed statistic T_{Cannes} to the bootstrap distribution using percentile confidence intervals and a two-sided empirical p value. A significant p value indicates that the observed level of comedy is too extreme to be attributed to random sampling alone.

We apply an almost identical procedure to the Cesar dataset. The main difference is that stratification is performed solely by year, since all films in both the Cesar selection and the corresponding TMDB comparison set are French. Regarding the temporal window, we restrict sampling to films released in the year preceding the award ceremony, in line with the official eligibility rules stating that eligible films must have been released the year before the ceremony. The same hierarchical fallback mechanism as described above is used to handle years that are underrepresented or absent in TMDB. By proceeding in this way, we approximate as closely as possible the institutional distributions of Cannes and the Cesar Awards and their respective eligibility criteria, while deliberately excluding genre information from the stratification, as genre is precisely the variable of interest in our analysis.
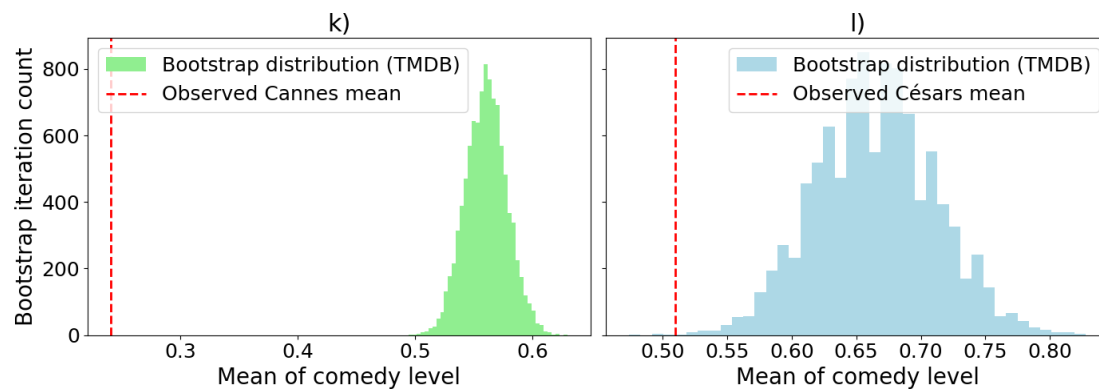
Each film is assigned an ordinal comedy score taking values 0, 1, or 2, where 0 indicates the absence of comedy, 1 corresponds to a mixed comedy, and 2 to a pure comedy. This score is derived from genre labels, based on the presence of the "comedy" tag and the possible co-occurrence of genres such as drama, romantic, action, or horror. The primary statistic of interest is the mean of this score within a given sample, which we interpret as the average intensity of comedy.

To assess how closely the bootstrap sampling scheme matches the target stratification, we report the frequency with which each sampling mode is used (exact match, country-window fallback, and year-only fallback). Because counting by strata may overrepresent cells with small sample sizes, we also report frequencies weighted by $n_{y,c}$, that is, the proportion of films actually drawn under each sampling mode.

2.3.3 Results

The observed mean comedy score for Cannes selections is far lower than the value expected under the bootstrap null distribution ($\mu_1 = 0.241$ and $\mu_2 = 0.561$, see **Figure 4.k**), lies well below the 95 percent bootstrap confidence interval [0.526, 0.596], and is therefore highly statistically significant (p<0.01), indicating a strong underrepresentation of comedy in the Festival de Cannes nominees relative to the TMDB comparison baseline.

The observed mean comedy score for César nominees (0.510) is substantially lower than the value expected under the bootstrap null distribution ($\mu_1 = 0.510$ and $\mu_2 = 0.665$, see **Figure 4.l**), falls well outside the 95 percent bootstrap confidence interval [0.575, 0.757], and is therefore statistically significant (p<0.01) as well, indicating a marked underrepresentation of comedy in the César Award nominees relative to the TMDB comparison baseline.

Figure 4 - Stratified bootstrap distributions of the mean comedy score for (k) the Cannes Film Festival selection and (l) the César Award selection, compared against their respective TMDB bootstrap baselines.

2.3.4 Conclusion

Although the winners of the top awards at the two major French film institutions do not exhibit a significantly lower level of comedy than the nominees, it could reasonably be argued that the number of observations is too small to reach statistical significance. The bias therefore lies elsewhere: nominees themselves display a significantly lower level of comedy than films that meet the eligibility criteria for selection.

That said, it is important to note that the TMDB dataset contains fewer films for the earliest years of these festivals than for later periods, which may raise concerns about the reliability of the results. One could argue that early years include only a limited number of films, likely skewed toward those that performed best commercially and, therefore, potentially toward comedies, which tend to perform better at the box office, as shown above.

Nevertheless, the consistency and robustness of the results provide meaningful insight into the selection biases exercised by juries in this type of competition, whether consciously or unconsciously.

These findings suggest that institutional bias against comedy does not manifest through differential treatment between winners and nominees, but rather through a more structural mechanism of exclusion. Comedy is significantly underrepresented among films selected for Cannes and the César Awards relative to the pool of eligible films, indicating that the primary filtering occurs upstream, at the stage of selection.

General Conclusion

This study set out to examine whether comedy is systematically devalued in contemporary cinema, and if so, to identify the stage at which this devaluation operates. By jointly analyzing box-office performance, audience and press ratings, award outcomes, and institutional selection processes, we sought to disentangle potential biases in reception from biases embedded in mechanisms of cultural consecration.

Our results first show that, when comedy is operationalized through genre labels, higher levels of comedy are associated with stronger box-office performance but lower evaluations from both critics and audiences. This pattern suggests that commercial success and symbolic valuation follow partially divergent logics, and that comedy may be particularly exposed to this tension. When comedy is instead inferred via a large language model, these associations become weaker and largely non-significant, with the exception of audience ratings and the strong alignment between LLM-based scores and genre labels. While this divergence may reflect limitations in statistical power or measurement reliability, it also raises the possibility that genre labels capture extrinsic expectations and market signals that shape evaluation independently of a film's intrinsic comedic intent.

Turning to institutional recognition, our analyses of the César Awards and the Festival de Cannes reveal no significant differences in the level of comedy between winners and nominees across major award categories. At face value, these results could suggest an absence of bias against comedy within award outcomes themselves. However, this interpretation changes fundamentally once the selection process is taken into account.

Using a stratified bootstrapping approach that compares festival selections to carefully constructed pools of eligible films, we find robust evidence of a systematic underrepresentation of comedy among nominees at both Cannes and the César Awards. Crucially, this bias emerges prior to award attribution: comedy is filtered out before films even enter the space in which symbolic distinctions are conferred. Once selected, comedic films do not appear to be penalized relative to non-comedic films in the competition for top prizes.

Taken together, these findings converge toward a central conclusion: the marginalization of comedy in prestigious film institutions does not primarily operate through differential treatment of already-selected films, but through an upstream mechanism of institutional gatekeeping. Any alleged disdain for comedy is therefore less a matter of overt evaluative bias at the final stage of consecration than a structural effect of selection criteria that implicitly favor certain aesthetic forms over others. Our preliminary findings also suggest a significant divergence between theatrical audiences and online reviewers, specifically within the comedy genre.

More broadly, this study highlights the importance of distinguishing between evaluation, consecration, and selection when analyzing cultural hierarchies. By showing that comedy's disadvantage lies chiefly at the level of nomination, our results contribute to a more precise understanding of how symbolic value is produced and distributed in the cinematic field. Future research could extend this framework to other genres, national contexts, or cultural institutions, and

further investigate the normative assumptions embedded in selection processes that shape the boundaries of legitimate art.

References

Chao, Brandon, et Ankit Sirmorya. "Automated Movie Genre Classification with LDA-based Topic Modeling. ». *International Journal of Computer Applications*, vol. 145, no. 13, juillet 2016, pp. 1-5.

Chu, Wei-Ta, et Hung-Jui Guo. "Movie Genre Classification based on Poster Images with Deep Neural Networks. ». *Proceedings of MUSA2'17*, Mountain View, CA, USA, 27 octobre 2017, pp. 39-45. ACM, doi:10.1145/3132515.3132516.

Fanny Camara and Nicolas Dupuis. Structural Estimation of Expert Strategic Bias: The Case of Movie Reviewers. Toulouse School of Economics Working Paper n° 14-534, 2014.

Xinyu (Elaine) Chen, Shuo Chen, and Param Vir Singh. Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products? Information Systems Research, 32(4):1173–1194, 2021.

Rachel Forshaw. Why Are Comedy Films So Critically Underrated? Master's thesis, University of Leeds, 2022.

Alexandra Georgakopoulou. On the Sociolinguistics of Popular Films: Funny Characters, Funny Voices. Journal of Modern Greek Studies, 18(1):119–133, 2000.

Matthew Kieran. Tragedy versus Comedy: On Why Comedy is the Equal of Tragedy. British Journal of Aesthetics, 53(1):1–15, 2013.

Ramona Mosse and Anna Street. Thinking Through Tragedy and Comedy: Some Provocations on Genre Matters. Performance Philosophy, 1(1):45–59, 2016.

Anna Street. Dramatic Measures: Comedy as Philosophical Paradigm. In Comedy and Critical Thought: Laughter as Resistance, edited by Iain MacKenzie and Anna Street, pp. 19–38. London: Rowman & Littlefield International, 2018.

Wang, Zhiqiang, et al. "Large Language Models Are Zero-Shot Text Classifiers". *arXiv*, 2 déc. 2023.