

## **First idea: Language patterns and multilingual films**

The dataset contains metadata about the languages spoken in each film, which allows us to explore patterns in the use of language across the global cinema. Is there a dominance of certain languages in the film industry? For example, is English dominant compared to other languages like Spanish or French? We can look at the percentage of films made in each language and see if the dominance of English-speaking films has decreased over time caused by more culturally diverse productions. Has the number of non-English films increased, reflecting a rise in international cinema? Also, we could examine multilingual films like how many films feature dialogue in more than one language. Is there a relationship between a film's genre and the likelihood of it being multilingual? For instance, action or adventure films may be more adaptable to multilingual audiences compared to genres like drama or romance. This analysis could reveal trends in language diversity and tell us if the globalization of cinema is embedding more inclusive productions in various languages.

## **Second idea: Predicting box office success based on metadata**

We can investigate whether factors like movie genre, runtime, release date, or language influence box office revenue. What are the most common predictors of a film's financial success? For instance, do action films generate more revenue than dramas? Does a film's runtime have an impact on its financial performance? We could explore whether movies released in the summer or during holiday seasons perform better than those released at other times of the year. By using machine learning models like linear regression, we can create a predictive model for box office success based on historical data. Is there a correlation between certain genres and higher revenue? Additionally, can we predict a new film's box office revenue based on its genre, language, runtime, and cast? Comparing the actual and predicted box office results, it could provide insight into the key factors that influence financial success in the film industry.

## **Third idea: Analyzing genre evolution over time**

The dataset provides metadata for over 42,000 films, including their release dates and genres, allowing us to track how the popularity of different genres has evolved over time. What are the trends in genre popularity over the past decades? For example, were Westerns more popular in the 1960s, and have superhero films surged in popularity in recent years? We can group films by decade and analyze the number of movies released in each genre during different time periods. Are there periods where certain genres, like sci-fi or horror, became more prevalent due to cultural or technological shifts? Also, we can explore how genre popularity differs across regions or countries. For instance, are action and adventure films more common in American cinema, while drama and romance are more popular in Europe? Using this data we can also analyze the impact of significant events—such as wars, economic crises, or technological advancements—on the trends of certain genres. For example, have war films increased in popularity during or after periods of conflict? Thank to visualizing these patterns, we can better understand the evolution of cinematic genres over time.