

Indian Movie Data Analysis

Executive summary

The analysis focuses on a curated movies dataset emphasizing Indian cinema, with Hindi, Tamil, and Malayalam comprising the largest language shares and India accounting for most titles in the corpus evaluated. Average runtimes trend upward in recent years while average ratings trend downward into 2025, and the catalogue is dominated by Drama-related genre combinations with notable contributions from prominent directors such as Mani Ratnam and Satyajit Ray. Among studios with high average ratings, Satyajit Ray Productions and Government of West Bengal stand out, though small sample sizes in several segments warrant cautious interpretation.

Data and scope

The dataset includes fields such as MovieTitle, ReleaseYear, RunTime, Genres, DirectorName, Rating, Studio, OriginCountry, OriginalLanguage, and cast, enabling exploration across time, geography, language, creators, and distribution entities. The corpus is highly India-centric by OriginCountry, with India contributing 664 movies while other countries appear in far smaller numbers, shaping the representativeness of many segment analyses. Languages are similarly concentrated, led by Hindi (318), Tamil (151), Malayalam (94), English (71), and Telugu (45), followed by smaller counts across regional and non-Indian languages.

Key findings: time trends

Average runtime by ReleaseYear increases into the 2020s, with the mean reaching approximately 148.43 minutes in 2025, suggesting longer films in the contemporary subset analyzed. Average rating by ReleaseYear exhibits variability but shows a recent decline, with the 2025 mean around 2.98, pointing either to a shift in catalogue composition or audience/reviewer dynamics in the latest entries.

Key findings: geography and language

By OriginCountry, the catalogue is dominated by India (664), with the next most frequent countries appearing at single- or low double-digit counts such as USA (10), France (9), Canada (7), and China (5) among others, indicating a strong India focus in the data source. Language distribution is led by Hindi (318), Tamil (151), Malayalam (94), English (71), and Telugu (45), with additional representation from Bengali/Bangla (16), Kannada (8), Marathi (3), and several single-title languages, shaping downstream genre and rating patterns.

Key findings: genres and creators

Genre combinations are numerous and skew toward Drama, with single-label “Drama” at 53 titles and “Drama, Romance” at 36, followed by “Romance, Drama” at 30 and “Drama, Comedy” at 21, showing that drama-adjacent storytelling dominates the corpus. Directors with the most titles include Mani Ratnam (15) and Satyajit Ray (13), followed by Shankar (9), Karthik Subbaraj (8), Anurag Kashyap (8), Karan Johar (8), Sanjay Leela Bhansali (8), Imtiaz Ali (7), and several others with smaller tallies, reflecting concentration among a set of prominent auteurs.

Key findings: studios and ratings

Among studios with high average ratings in the summary, Satyajit Ray Productions shows about 4.36 and Government of West Bengal around 4.32, with others such as Tribal Art Productions (4.26) also performing strongly in their respective subsets. Because studio-level averages can be sensitive to small sample sizes, these figures should be interpreted alongside the number of titles per studio to avoid over-weighting outliers.