

Proposal: Analysis of MovieLens Dataset

Datasets

1. **movies_metadata.csv**: Information on 45,000 movies including posters, backdrops, budget, revenue, release dates, languages, production countries, and companies.
2. **keywords.csv**: Plot keywords for movies, available as a stringified JSON Object.
3. **credits.csv**: Cast and crew information, available as a stringified JSON Object.
4. **links.csv**: TMDb and IMDb IDs of all movies in the Full MovieLens dataset.
5. **links_small.csv**: TMDb and IMDb IDs for a subset of 9,000 movies.
6. **ratings_small.csv**: Subset of 100,000 ratings from 700 users on 9,000 movies.

Proposed Questions

1. What are the most common genres among the top-rated movies?
2. How do budget and revenue correlate with movie ratings?
3. Which keywords are most associated with high-grossing movies?
4. What is the distribution of movie release dates, and are there any trends over time?
5. How do cast and crew attributes (e.g., director, lead actors) affect movie ratings?
6. What are the differences in user ratings distribution across different genres?
7. How do movie ratings vary by production countries?
8. What trends can be observed in movie budgets and revenues over time?
9. How effective are keywords in predicting movie genres?
10. What are the characteristics of movies that receive the most ratings?

Goals

1. **Analyze Genre Popularity**: Determine the most common genres among the top-rated movies to understand audience preferences.
2. **Correlation Analysis**: Investigate the relationship between movie budgets, revenues, and ratings to identify potential patterns.
3. **Keyword Impact**: Examine the influence of plot keywords on a movie's financial success and popularity.
4. **Release Date Trends**: Analyze the distribution and trends of movie release dates to identify peak periods for releases.
5. **Cast and Crew Influence**: Assess how the attributes of cast and crew members correlate with movie ratings to understand the impact of key personnel.
6. **Ratings Distribution by Genre**: Explore how user ratings vary across different genres to identify which genres receive the most positive feedback.
7. **Country-wise Ratings**: Investigate how movie ratings differ by production countries to understand geographic influences on ratings.
8. **Budget and Revenue Trends**: Analyze trends in movie budgets and revenues over time to identify patterns and anomalies.

9. **Predictive Power of Keywords:** Evaluate the effectiveness of keywords in predicting movie genres to improve recommendation systems.
10. **Characteristics of Highly Rated Movies:** Identify the common characteristics of movies that receive the most ratings to inform content creation and marketing strategies.