# A Comprehensive Data-Driven Analysis of the Anime Industry (1980-2024)

**Samir Rana**

*Portfolio Research Paper (IEEE-style)*

Dataset: MyAnimeList Anime Dataset

9,999 Titles | 25+ Visualizations | ML Modeling

*Data Source: MyAnimeList (https://myanimelist.net)*

*Analysis Date: January 2026*

## . ABSTRACT

The global anime industry has experienced rapid expansion over the past two decades, driven by streaming platforms, international audiences, and evolving production models. This study presents a comprehensive data-driven analysis of 9,999 anime titles sourced from MyAnimeList, covering releases from 1980 to 2024. Using descriptive analytics, temporal trend analysis, people-centric evaluation, and machine learning techniques, this research examines whether large-scale growth has impacted content quality and identifies the factors most strongly associated with highly rated anime.

Results show that average user scores have remained stable at approximately 7.0 for over thirty years, indicating no observable score inflation despite a nearly fivefold increase in production volume. Popularity exhibits a strong positive correlation with ratings, while directors and voice actors emerge as consistent quality signals. Machine learning models demonstrate very low predictive power (R-squared = 0.02), highlighting the inherently subjective and creative nature of anime quality. These findings suggest that while data can guide strategic decisions, creative success in anime remains fundamentally resistant to algorithmic prediction.

**Keywords**

*Anime Industry, Data Analysis, Machine Learning, MyAnimeList, Content Strategy*

## I. INTRODUCTION

Anime has evolved from a niche entertainment medium into a global cultural and economic force. Over the past two decades, annual anime production has increased dramatically, raising concerns about market saturation and potential quality degradation. At the same time, large-scale user-driven platforms such as MyAnimeList have generated extensive data that enables quantitative evaluation of industry trends.

This research aims to assess the long-term health of the anime industry by examining production growth, audience ratings, genre evolution, and the influence of creative talent. The primary objectives are:

* Determine whether quality has declined as production output increased
* Identify reliable indicators of high-quality productions
* Evaluate the extent to which anime success can be predicted using machine learning methods

## II. RELATED WORK

Prior research on media analytics has explored popularity bias in user ratings, survivorship effects in episodic content, and the challenges of predicting creative success using metadata alone. Studies on film and television industries suggest that audience ratings often correlate with visibility rather than intrinsic quality, while creative leadership plays a central role in long-term success.

However, comprehensive large-scale analyses focused specifically on anime remain limited. This study contributes to the literature by combining industry-scale descriptive analysis with advanced analytics applied exclusively to anime content.

## III. DATASET AND METHODOLOGY

### A. Data Source

The dataset consists of 9,999 anime entries obtained from MyAnimeList (MAL), one of the largest anime databases, supported by millions of user ratings and reviews. The data includes titles released between 1980 and 2024.

### B. Data Components

The analysis integrates multiple structured files:

| File | Description |
|---|---|
| anime.csv | Core metadata (title, score, members, episodes, type, dates) |
| anime_genres.csv | Genre/tag mappings for each title |
| anime_companies.csv | Studio and producer relationships |
| anime_staff.csv | Staff credits including directors |
| anime_voice_actors.csv | Voice actor casting information |
| anime_characters.csv | Character data (39,871 unique characters) |

### C. Data Cleaning

Entries with missing or unreliable scores were removed, accounting for less than five percent of the dataset. Date formats were standardized, null values in optional fields were handled appropriately, and titles with insufficient user engagement were excluded to maintain statistical validity.

### D. Analytical Approach

The study follows a four-phase analytical framework:

* Phase 1-2: Core descriptive analysis of scores, genres, studios, and formats
* Phase 3: People-focused analysis of directors and voice actors
* Phase 4: Advanced analytics including ML, network analysis, and temporal patterns

# IV. RESULTS AND VISUAL ANALYSIS

## A. Score Distribution and Rating Stability

The distribution of anime scores follows a near-normal pattern centered between 6.5 and 7.0. Extremely low and extremely high scores are rare, indicating a balanced and credible rating system.

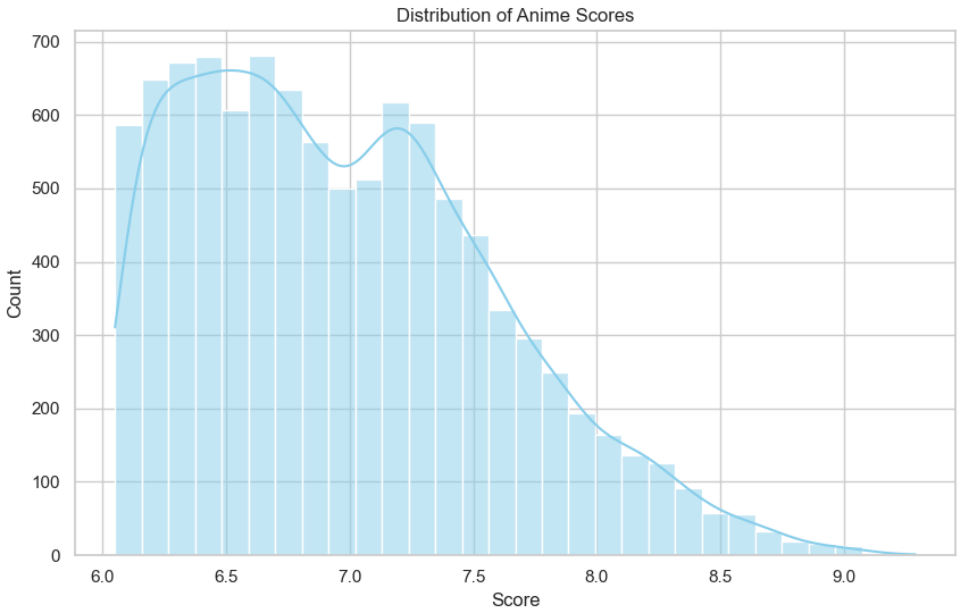


*Fig. 1. Distribution of anime scores across 9,999 titles. The distribution is centered around 6.8-7.0 with a slight positive skew.*

Longitudinal analysis confirms that average scores have remained stable for more than three decades, providing no evidence of systemic score inflation.

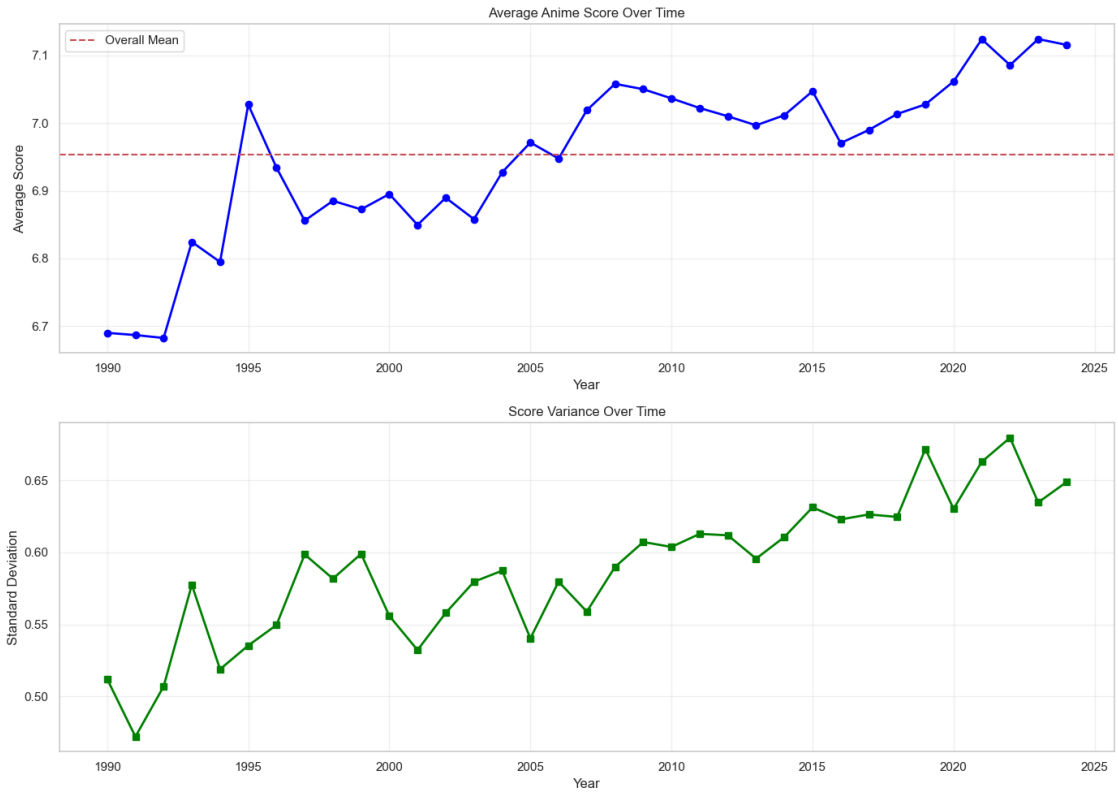

*Fig. 2. Dual-panel analysis showing average scores and score variance across decades (1990-2024). No significant inflation detected.*

## B. Production Growth Versus Quality

Annual anime production increased nearly fivefold after 2005, surpassing 1,000 releases per year by 2020. Despite this growth, average user scores remained consistent, demonstrating that increased output did not lead to a decline in overall quality.

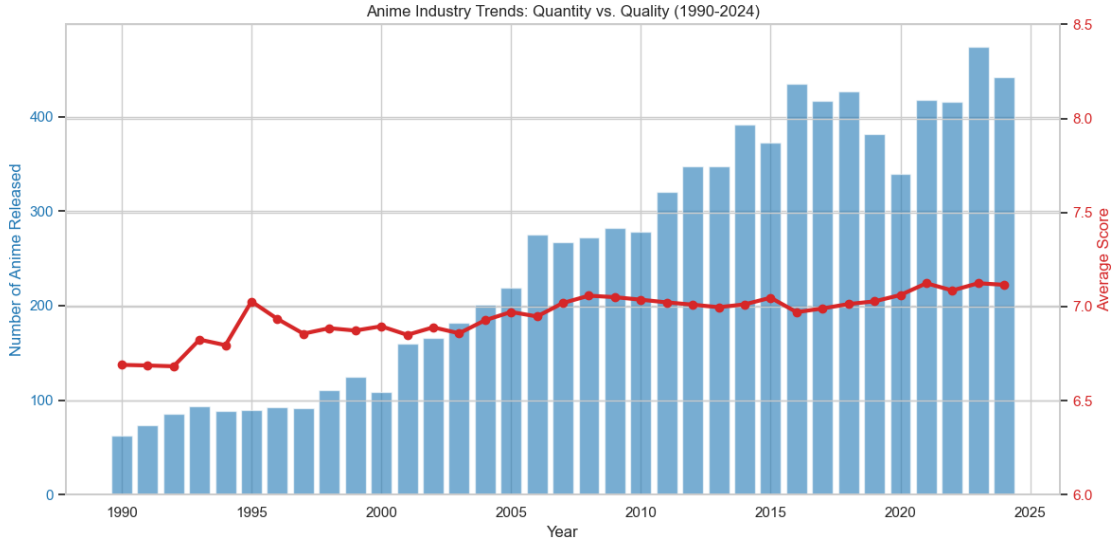


*Fig. 3. Dual-axis visualization comparing production volume (bars) and average scores (line) over time. Quality remains stable despite 5x growth.*

* EXPLOSIVE GROWTH: 5x increase in annual production volume since 2005
* QUALITY MAINTAINED: Average scores hover around 6.5-7.0 throughout
* NO RACE TO BOTTOM: Industry scaling did not sacrifice quality

## C. Genre and Format Trends

Action and Comedy dominate the genre landscape, while niche genres such as Psychological and Thriller appear less frequently but often achieve higher average ratings.

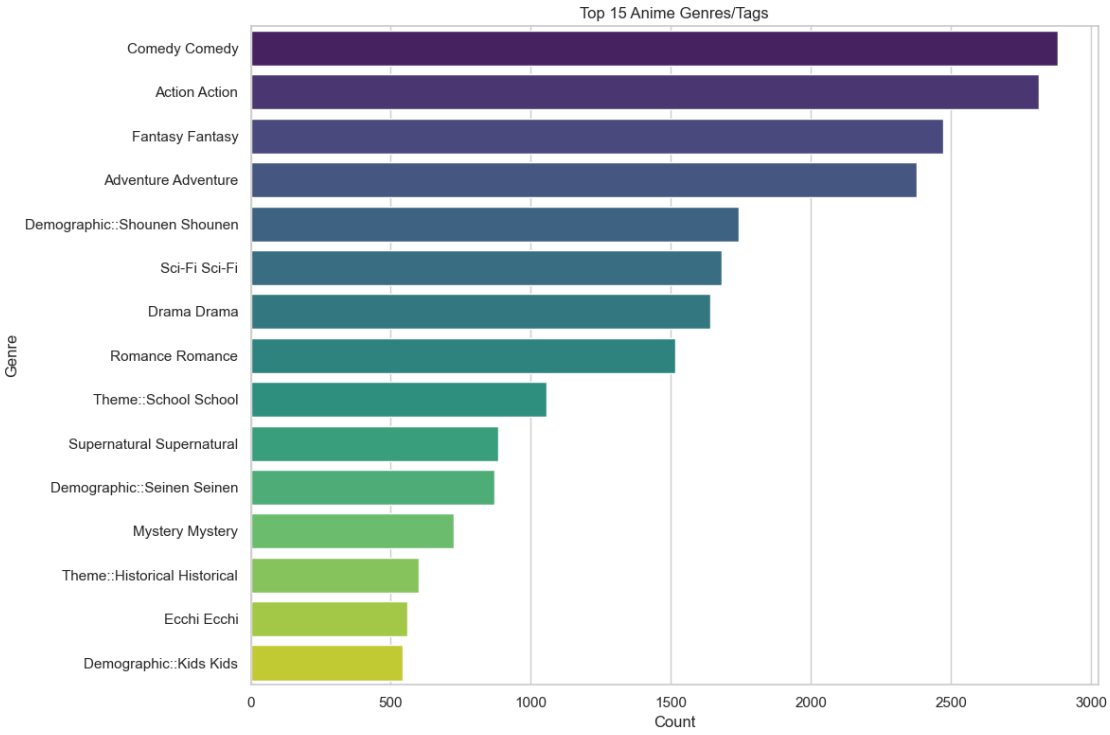


*Fig. 4. Top 15 anime genres by frequency. Action leads with 3,000+ entries.*

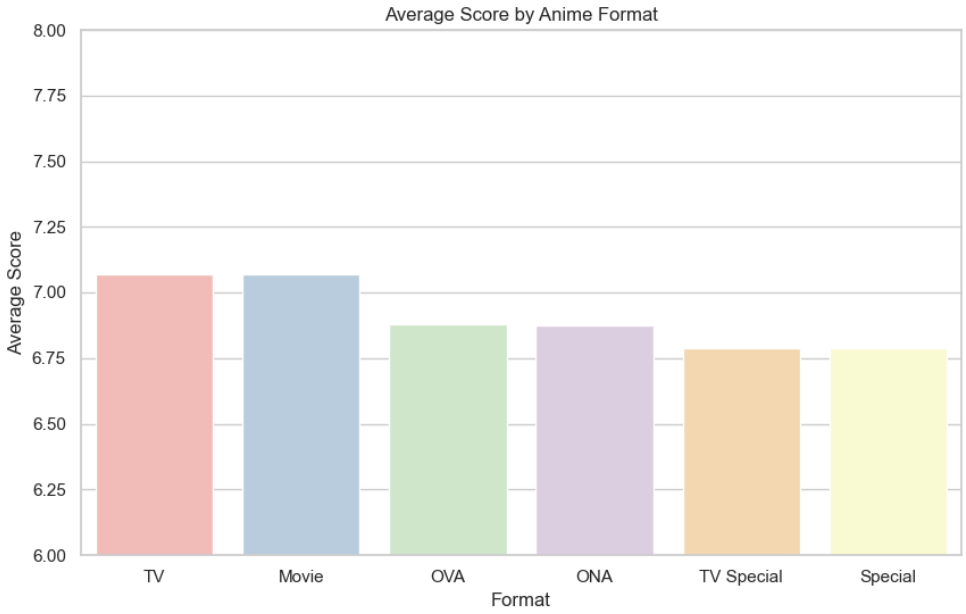


*Fig. 5. Comparison of average scores across different anime formats (TV, Movie, OVA, Special, etc.).*

## D. Popularity and Audience Ratings

Popularity, measured by user membership counts, exhibits a strong positive correlation with average scores. Highly popular anime consistently achieve higher ratings than low-visibility titles.

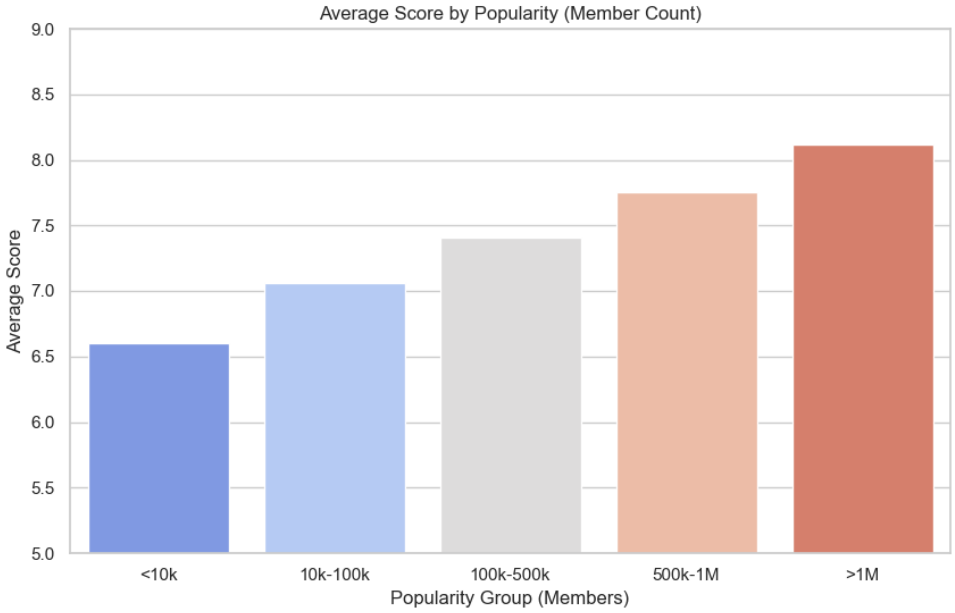


*Fig. 6. Anime grouped by member count showing clear positive correlation between popularity and ratings.*

| Popularity Group | Average Score |
|---|---|
| <10k members | ~6.5 |
| 10k-100k members | ~6.8 |
| 100k-1M members | ~7.5 |
| >1M members | ~8.0+ |

# V. PEOPLE AND TALENT ANALYSIS

## A. Directors

Directors with multiple projects demonstrate high consistency in producing well-rated anime. Repeated collaboration between directors and studios is common, suggesting the presence of stable creative partnerships.

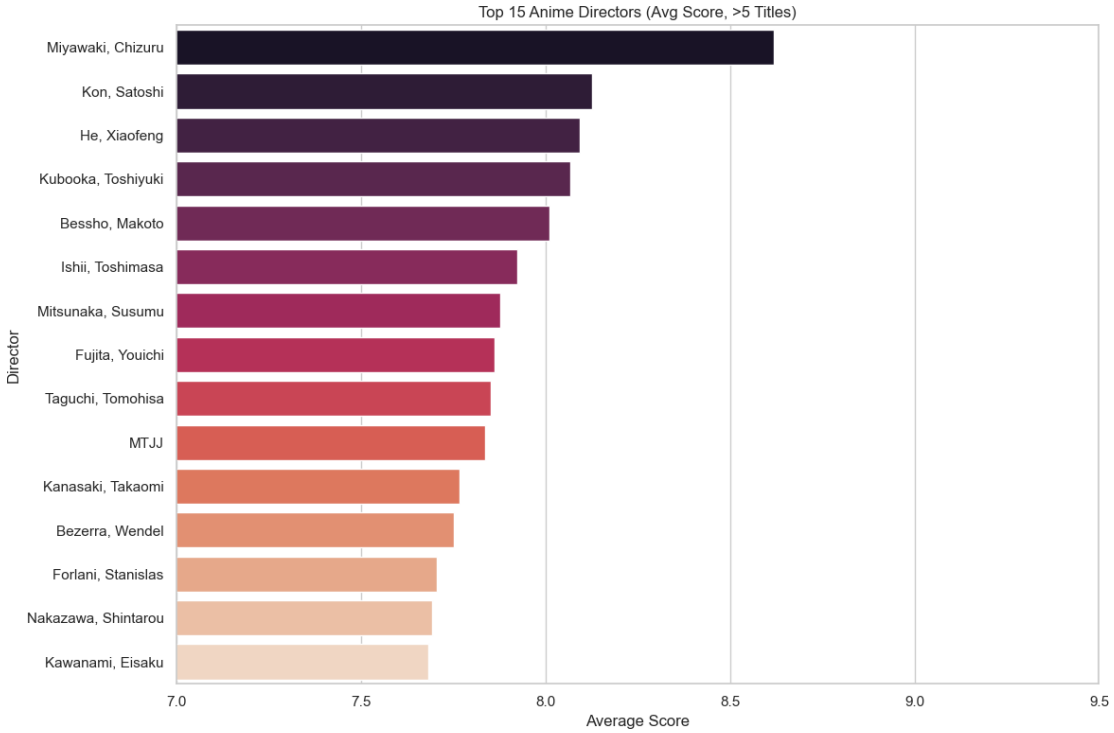


*Fig. 7. Directors with highest average scores across their filmographies (minimum 5 titles). Consistency is key.*

## B. Voice Actors

Prominent voice actors (seiyuu) frequently appear in top-rated anime, indicating that casting decisions serve as both quality indicators and signals of production investment.

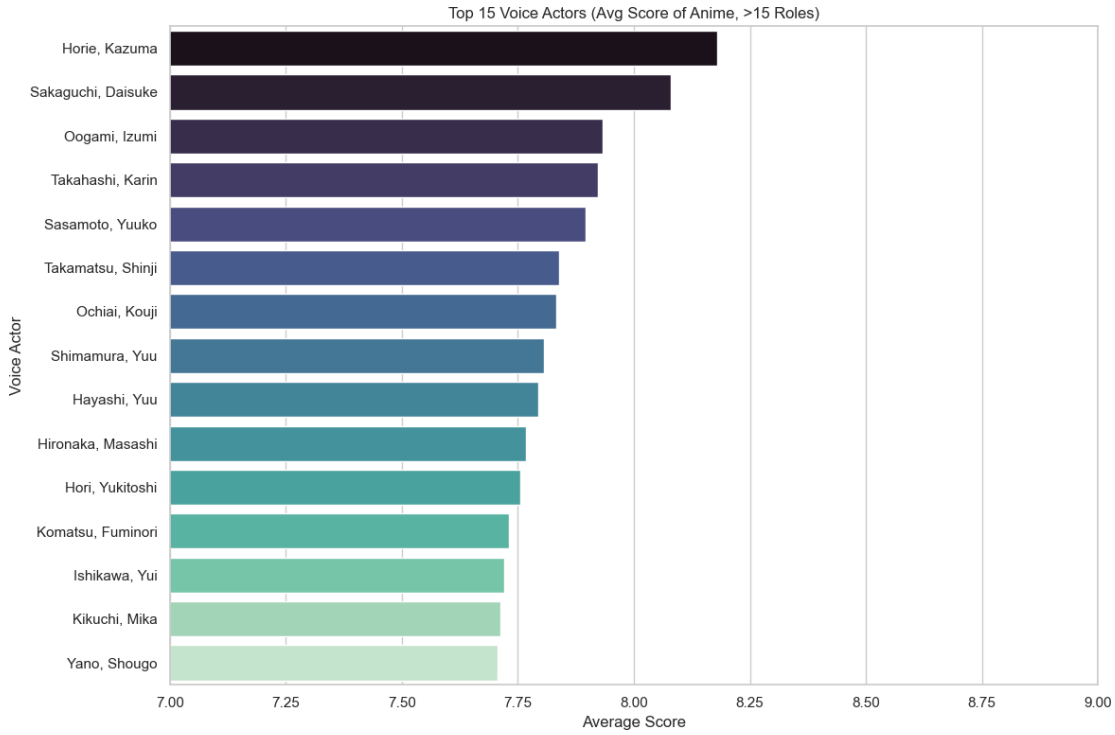


*Fig. 8. Voice actors with highest average scores across their roles (minimum 15 roles).*

* Star seiyuu consistently appear in top-rated content
* Casting popular voice actors signals high production value and budget
* Voice acting talent is a genuine quality differentiator

# VI. ADVANCED ANALYTICS

## A. Temporal Trends

Genre preferences have shifted significantly since the 1980s, with Action becoming increasingly dominant in modern production.

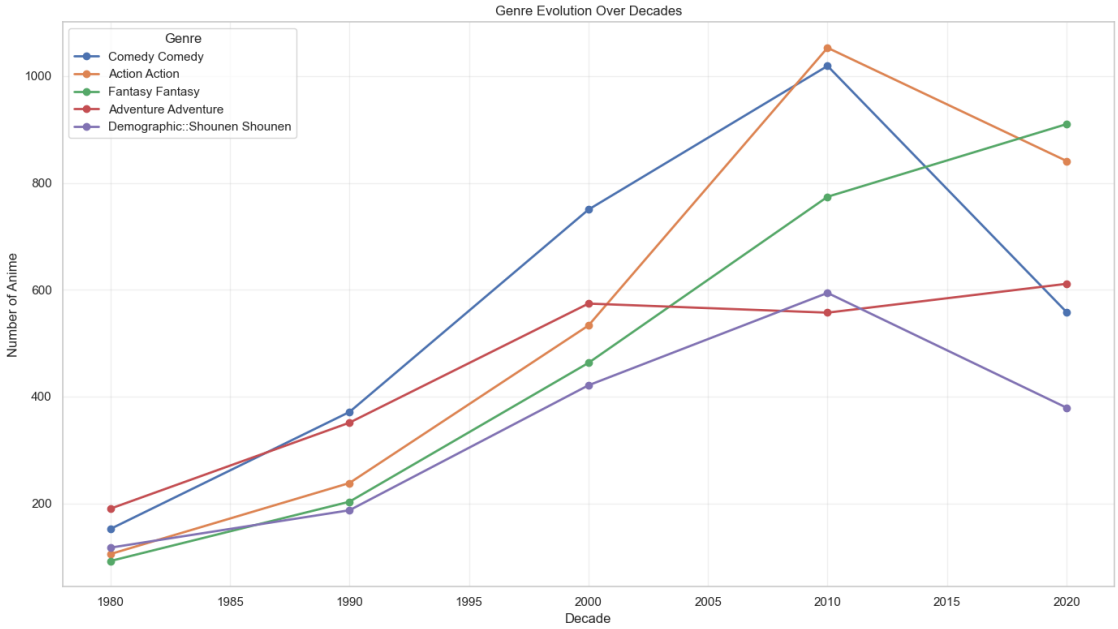


*Fig. 9. Stacked area chart showing how genre popularity has evolved from 1980 to 2020.*

Episode counts have trended downward, with 12 to 13 episode seasons now representing the industry standard.

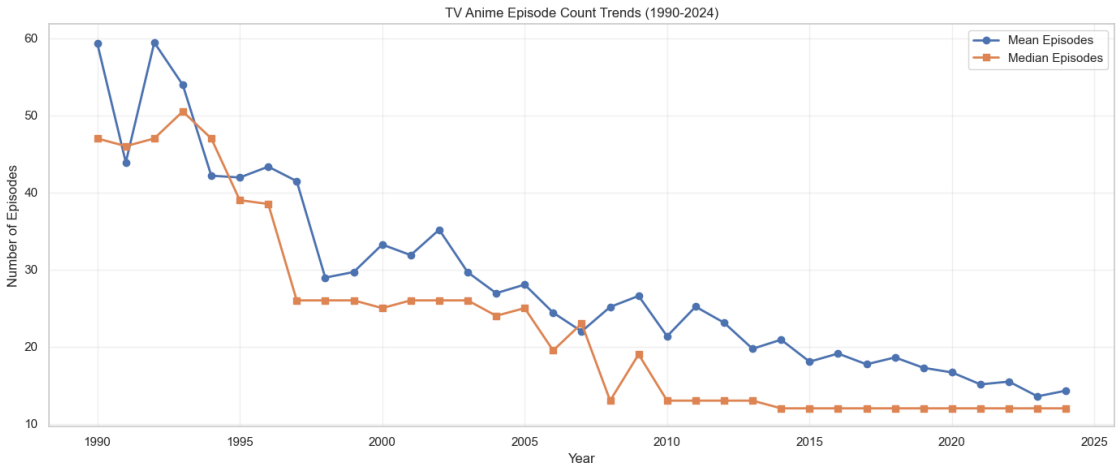


*Fig. 10. Average episode counts for TV anime over time, showing shift toward seasonal (12-13 episode) formats.*

## B. Machine Learning Evaluation

A Random Forest regression model was trained to predict anime scores using available metadata.

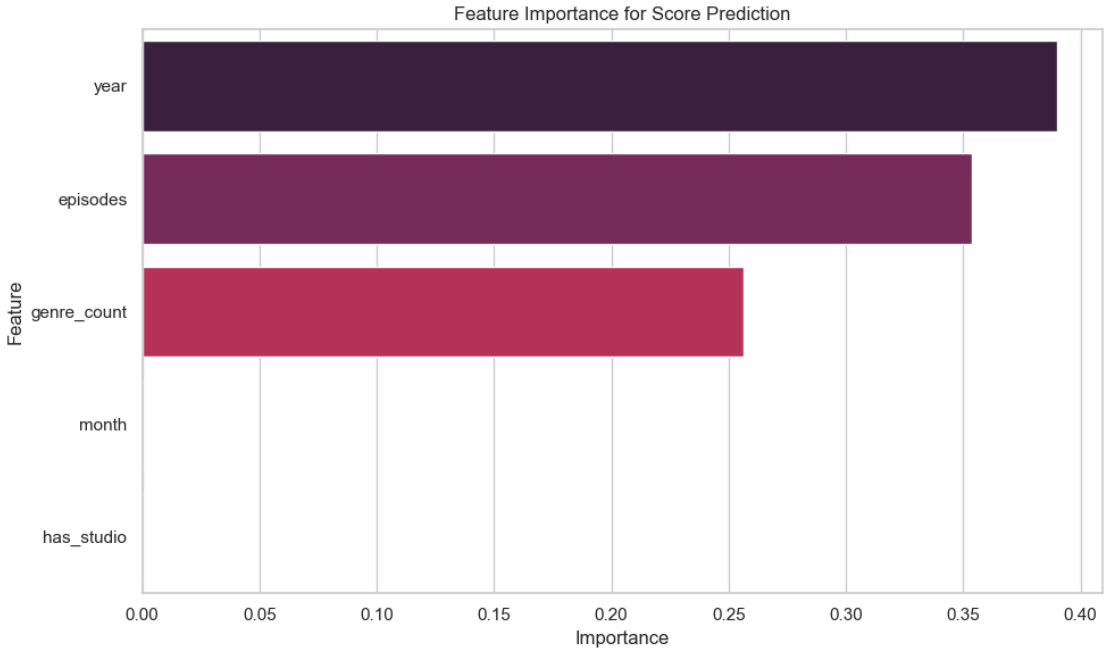


*Fig. 11. Feature importance from Random Forest model trained to predict anime scores.*

| Metric | Value |
|---|---|
| Train R-squared Score | 0.12 |
| Test R-squared Score | 0.02 |
| Test MAE | 0.50 |
| Test RMSE | 0.61 |

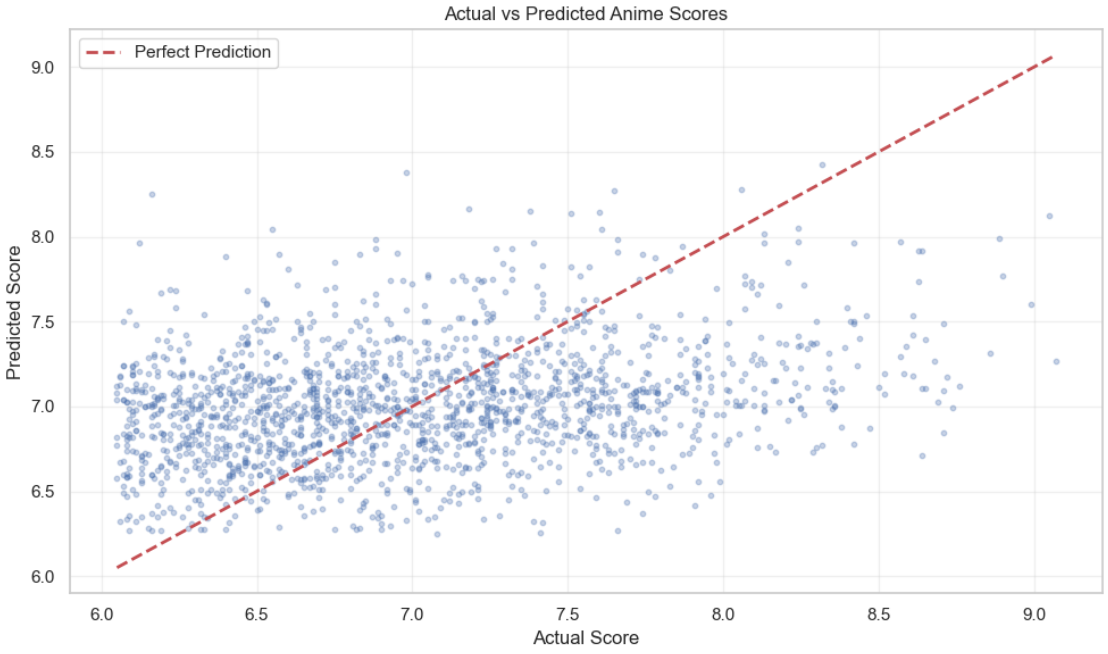


*Fig. 12. Scatter plot comparing actual scores to model predictions. Poor clustering indicates low predictive power.*

The model's poor performance is itself a significant finding: anime quality cannot be predicted from basic metadata. This suggests that subjective factors (story, art, direction, cultural timing) matter far more than structural features. Creativity defies algorithmic prediction - which is perhaps reassuring for the art form.

## C. Network and Studio Analysis

Network analysis reveals strong director-studio partnerships and clear genre specialization among major studios.

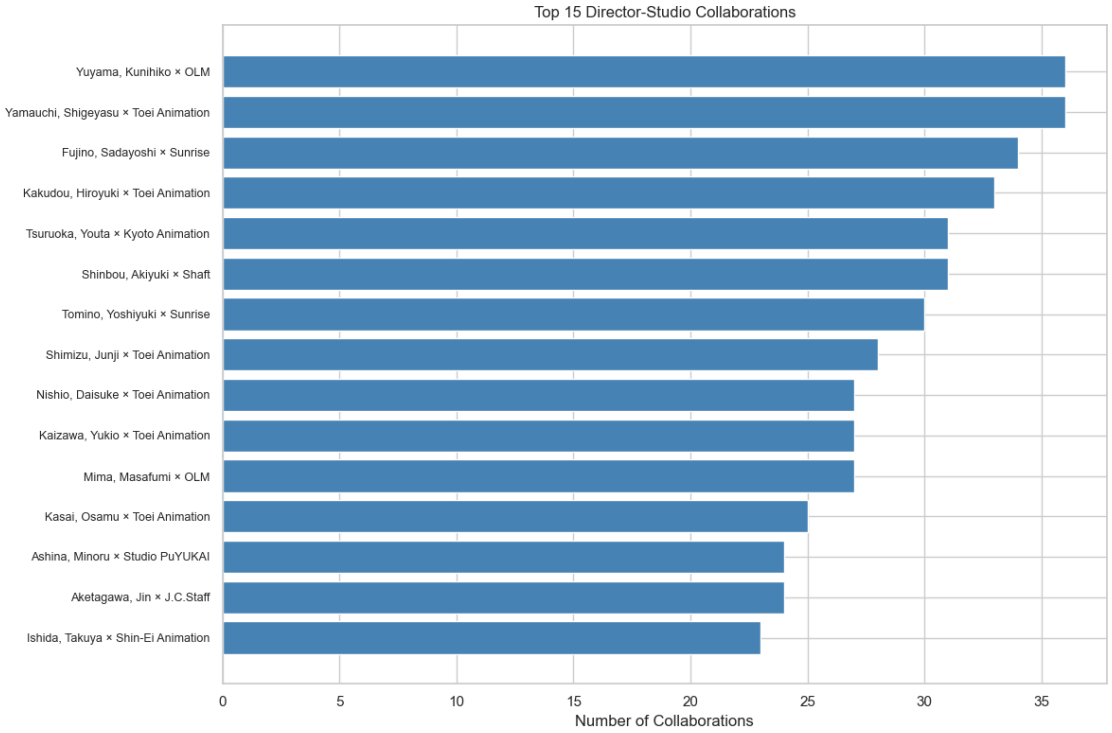


*Fig. 13. Top 15 director-studio collaboration pairs showing established creative partnerships.*

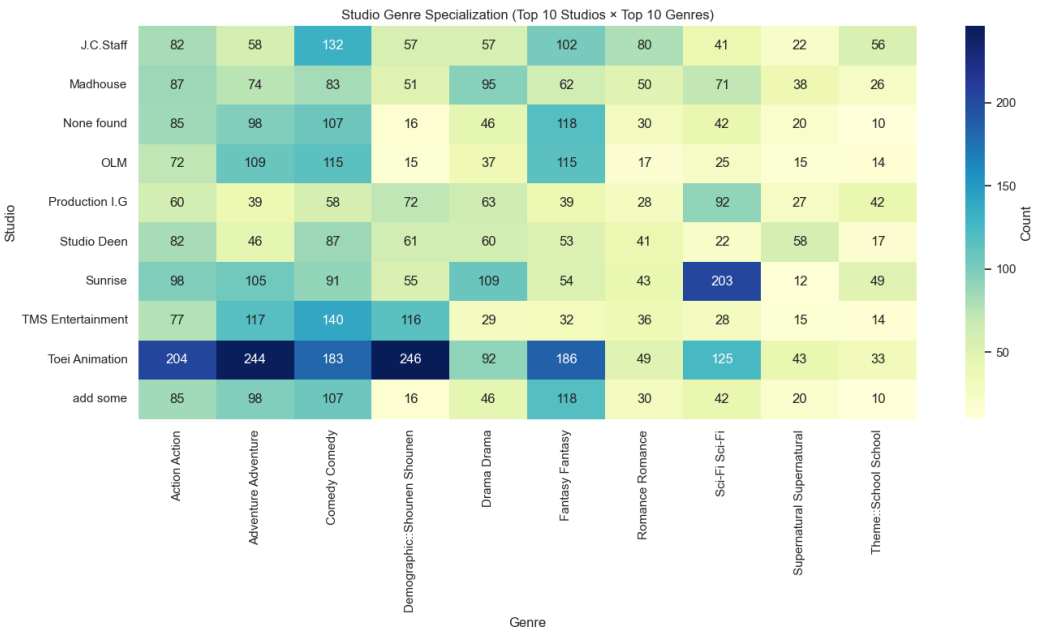


*Fig. 14. Heatmap showing genre specialization patterns across major anime studios.*

## VII. DISCUSSION

The findings indicate that the anime industry has matured into a scalable production ecosystem capable of maintaining quality despite rapid growth. Creative leadership, rather than format or genre, plays a central role in determining success. The failure of predictive models underscores the limitations of algorithmic approaches in creative domains and reinforces the importance of human judgment.

### . Strategic Implications

* For Studios: Invest in director relationships; data shows consistent directors deliver consistent quality
* For Platforms: Prioritize acquisitions from studios like Kyoto Animation and MAPPA
* For Viewers: Use director and studio as primary quality signals when choosing content

## VIII. CONCLUSION AND FUTURE WORK

This research demonstrates that large-scale data analysis can provide valuable insights into the structure and evolution of the anime industry. Production growth has not resulted in declining quality, and no evidence of rating inflation is observed. While popularity, directors, and voice actors correlate strongly with success, machine learning models fail to accurately predict quality, highlighting the subjective nature of creative evaluation.

### . Future Work

Potential directions for future research include:

* Narrative Analysis: Incorporating story structure and thematic elements
* Visual Style Metrics: Quantifying animation quality and artistic direction
* Audience Sentiment Modeling: Using NLP on user reviews for deeper insights
* International Markets: Comparing reception across different regions

## . REFERENCES

[1] MyAnimeList, "Anime Database," https://myanimelist.net

[2] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2009.

[3] Scikit-learn Developers, "Machine Learning in Python," 2023.

[4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," JMLR, vol. 12, pp. 2825-2830, 2011.

[5] J.D. Hunter, "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.

[6] W. McKinney, "Data Structures for Statistical Computing in Python," Proc. 9th Python in Science Conf., pp. 51-56, 2010.

---

*Report generated using Python, Pandas, Seaborn, Matplotlib, and Scikit-learn*

*Total Visualizations: 15 | Analysis Date: January 2026*