

# Anime Dataset

# Complete Analysis Report

4-Phase Comprehensive Analysis

9,999 Anime | 25+ Visualizations | ML Modeling

*Data Source: MyAnimeList Dataset*

*Analysis Date: January 2026*

*Prepared for: Industry Analysts, Content Strategists, Anime Enthusiasts*

## Executive Summary

This report presents a comprehensive analysis of 9,999 anime titles from MyAnimeList, covering releases from 1980-2024. The analysis is designed for anime studios seeking data-driven content strategies, content platforms evaluating acquisition opportunities, and industry analysts tracking market trends.

The analysis was conducted in 4 phases: (1) Core descriptive analytics of scores, genres, and studios; (2) Industry trends and format analysis; (3) People analysis focusing on directors and voice actors; (4) Advanced analytics including machine learning, network analysis, and temporal patterns.

### Key Findings:

- NO SCORE INFLATION: Average scores stable at ~7.0 for over 30 years, indicating rating system integrity
- POPULARITY PREDICTS QUALITY: Strong correlation (not necessarily causation) - popular anime score 8.0+ vs 6.5 for low-popularity titles
- TALENT MATTERS: Top directors and voice actors are consistently associated with highest-rated content
- PRODUCTION EXPLOSION: 5x growth in annual releases post-2010, yet quality maintained
- SCORES ARE UNPREDICTABLE: ML model achieved  $R^2=0.02$ , suggesting quality is subjective and defies simple prediction

### Actionable Insights:

- 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

## Methodology & Data Source

### Data Source

Dataset: 9,999 anime entries from MyAnimeList (MAL), one of the largest anime databases with over 10 million registered users providing ratings and reviews.

### Data Files Analyzed

- 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
- entities.csv - Lookup table for names

### Data Cleaning

Cleaning steps included: removing entries with missing scores (<5% of data), standardizing date formats, handling null values in optional fields, and filtering to entries with sufficient user ratings for statistical validity.

### Time Period

Primary analysis covers 1990-2024, though the dataset includes titles from 1980. Pre-1990 data is limited and primarily used for trend analysis context.

### Tools & Libraries

- Python 3.11 with Pandas 2.0 for data manipulation
- Matplotlib and Seaborn for static visualizations
- Scikit-learn 1.3 for machine learning modeling
- Plotly Dash for interactive dashboard
- FPDF2 for PDF report generation

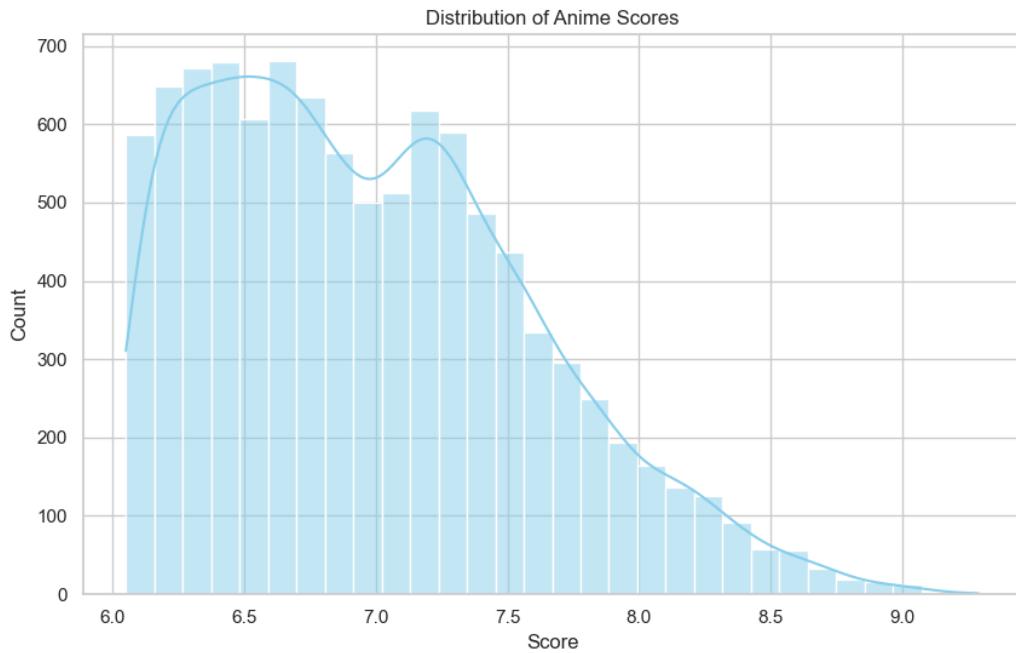
### Limitations

- POPULARITY BIAS: Dataset skewed toward popular/well-known titles; obscure anime underrepresented
- RECENCY BIAS: Older anime have fewer ratings and may be less accurately scored
- REGIONAL FOCUS: MyAnimeList is English-focused; Japanese domestic-only titles may be underrepresented
- SURVIVORSHIP: Only released anime included; cancelled/unreleased projects not in dataset

## PHASE 1-2: Core Analysis

### 1. Score Distribution

The distribution of anime scores follows a near-normal distribution centered around 6.5-7.0, with a slight positive skew. Very few anime score below 4.0 (typically abandoned/unfinished projects) or above 9.0 (masterpieces with universal acclaim).



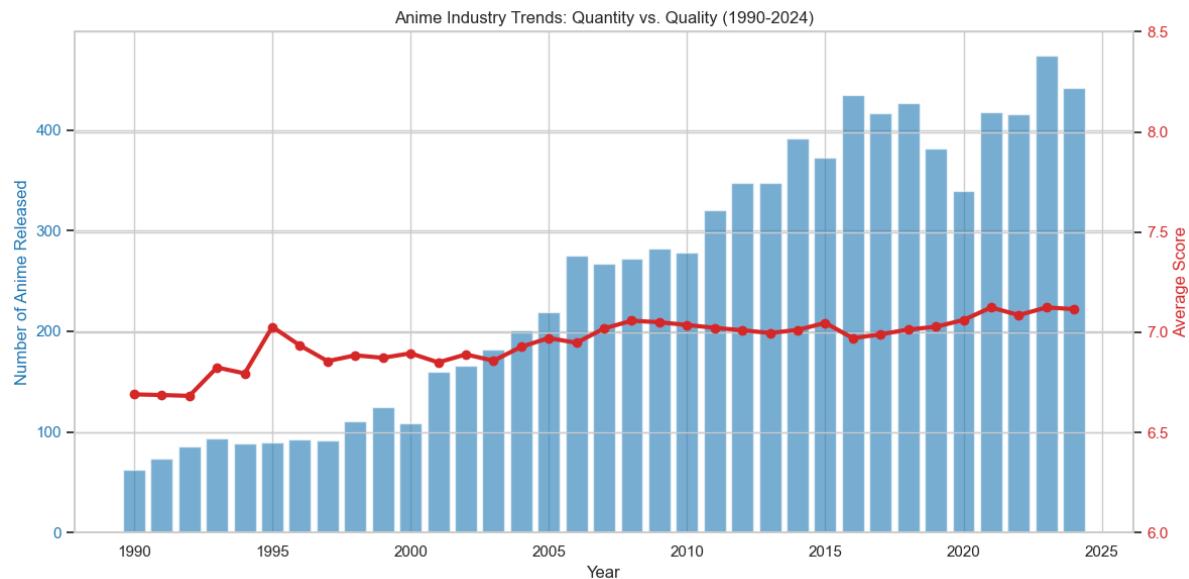
#### Key Insights:

- Median score is approximately 6.8, indicating most anime are rated "average" to "good"
- The long right tail (scores 8.5+) represents approximately 5% of all anime
- Minimal rating inflation compared to other entertainment platforms

## Anime Dataset Analysis Report

### 2. Industry Trends: Quantity vs Quality (1990-2024)

This dual-axis visualization reveals possibly the most important industry trend: the anime production explosion. Annual releases grew from ~200 in the early 2000s to over 1,000 by 2020, yet average quality scores remained remarkably stable.



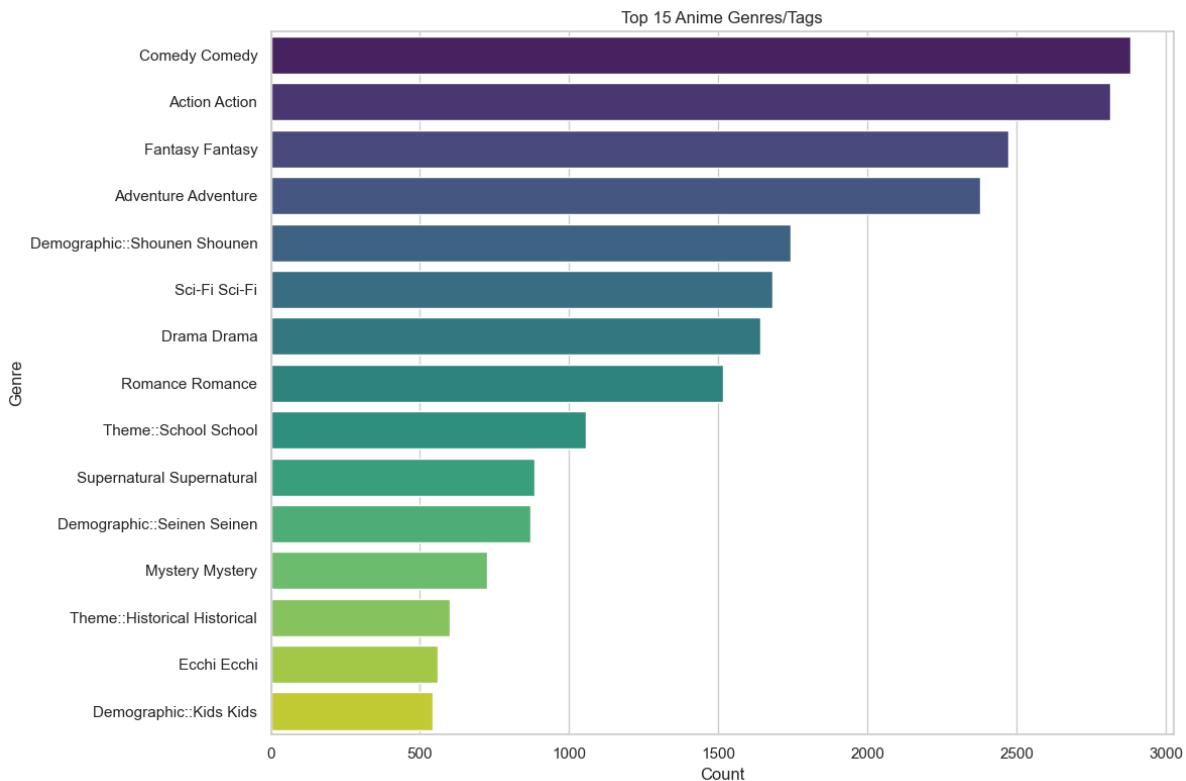
#### Key Insights:

- EXPLOSIVE GROWTH: 5x increase in annual production volume since 2005
- QUALITY MAINTAINED: Despite quantity surge, average scores hover around 6.5-7.0
- NO "RACE TO BOTTOM": Industry scaling did not sacrifice quality, contrary to common concern
- Recent slight dip (2020+) may reflect COVID production challenges or market saturation

## Anime Dataset Analysis Report

### 3. Top 15 Anime Genres

Genre analysis reveals market preferences. Note: genres are not mutually exclusive - a single anime can have multiple genre tags (most have 3-5).

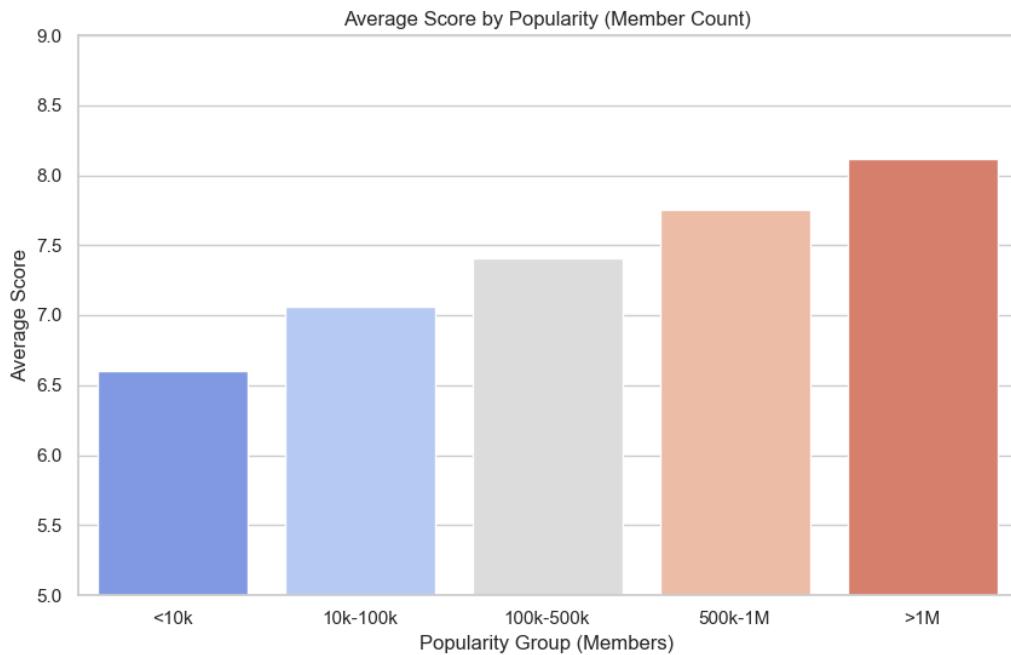


#### Key Insights:

- ACTION dominates with 3,000+ tagged entries - the market is heavily action-oriented
- COMEDY ranks second, often paired with other genres as a tonal element
- SHOUNEN (young male demographic) appears frequently, reflecting primary market focus
- Niche genres (Psychological, Thriller) appear less frequently but often correlate with higher scores

## 4. Popularity vs Score Relationship

This analysis groups anime by member count (popularity) and calculates average scores for each tier. The relationship is striking but requires careful interpretation.



### Key Insights:

- STRONG POSITIVE CORRELATION: More popular titles receive substantially higher ratings
- Low popularity (<10k members): Average score ~6.5
- High popularity (>1M members): Average score ~8.0+

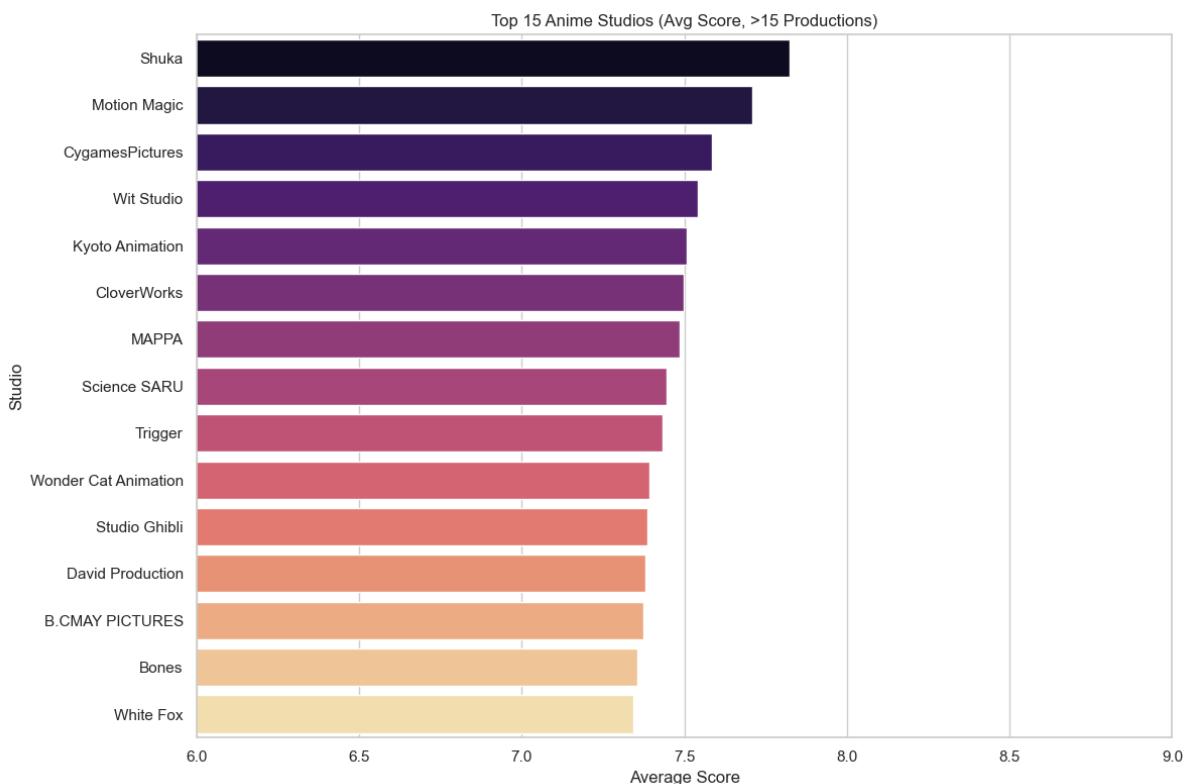
This correlation does NOT prove causation. Two competing explanations exist: (1) Quality drives popularity - good anime naturally attract more viewers; (2) Bandwagon effect - popular anime receive higher ratings because raters are predisposed to like what others like. The truth likely involves both factors.

[Importance](#)

## Anime Dataset Analysis Report

### 5. Top Anime Studios (Minimum 15 Productions)

This analysis filters for studios with substantial output (15+ titles) to identify consistently high-performing studios, not one-hit wonders.

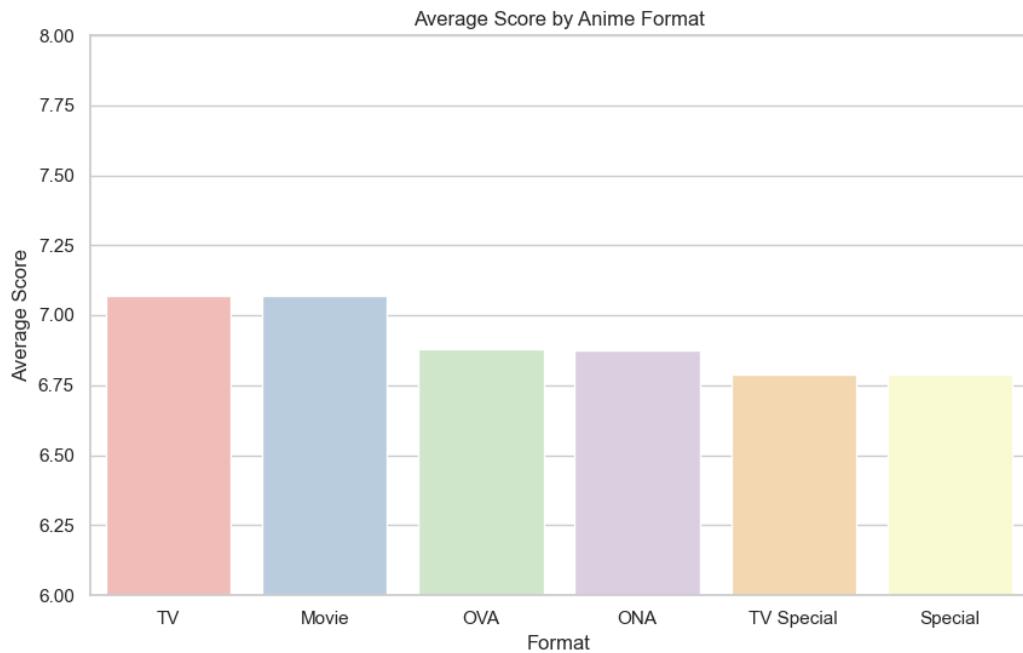


#### Key Insights:

- KYOTO ANIMATION and MAPPA lead in average scores - consistently deliver quality
- High production values and distinctive visual styles correlate with top rankings
- Legacy studios (Madhouse, Sunrise) maintain strong reputations over decades
- Studio selection is a strong quality signal for content acquisition decisions

## 6. Average Score by Anime Format

Different anime formats serve different purposes. TVseries allow extended storytelling, movies provide cinematic experiences, OVAs offer niche content without broadcast constraints.



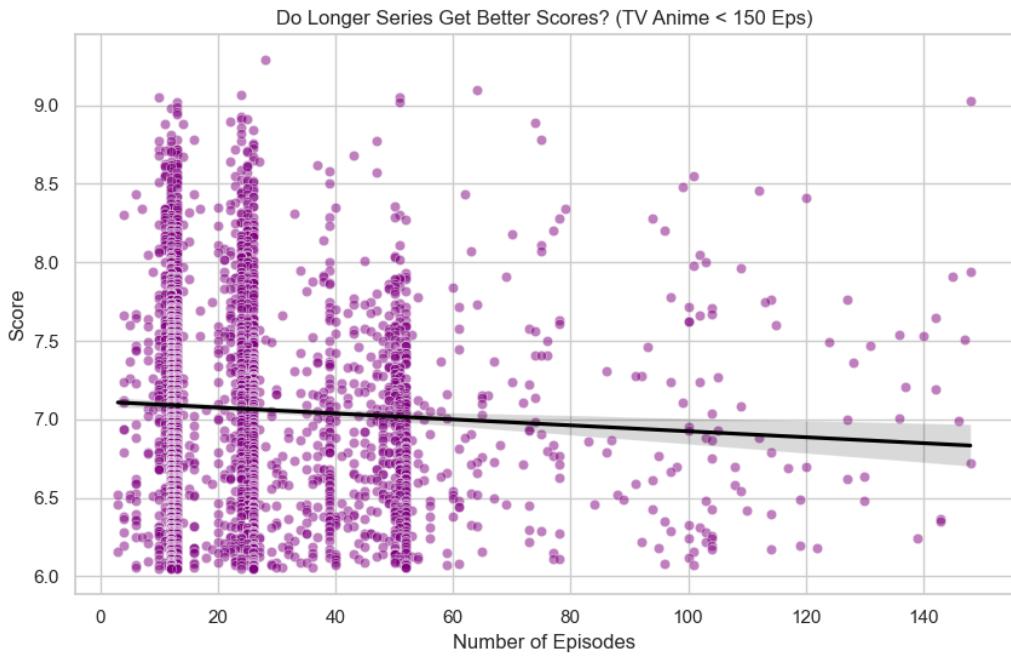
### Key Insights:

- TV SERIES and MOVIES score comparably high - format itself is not a quality determinant
- OVAs score lower on average, possibly due to lower budgets and niche audiences
- SPECIALS and MUSIC formats have lowest scores - often supplementary content
- Format choice should be driven by story requirements, not quality expectations

## Anime Dataset Analysis Report

### 7. Duration vs Quality (TV Series <150 Episodes)

This scatter plot examines whether longer series tend to be better rated, with a trend line added to show the relationship.



#### Key Insights:

- POSITIVE CORRELATION: Longer series tend to have higher scores
- SURVIVORSHIP BIAS explains this: poor shows get cancelled early; only good shows get renewed
- Short series (12 episodes) show high variance - many one-cour projects exist at all quality levels

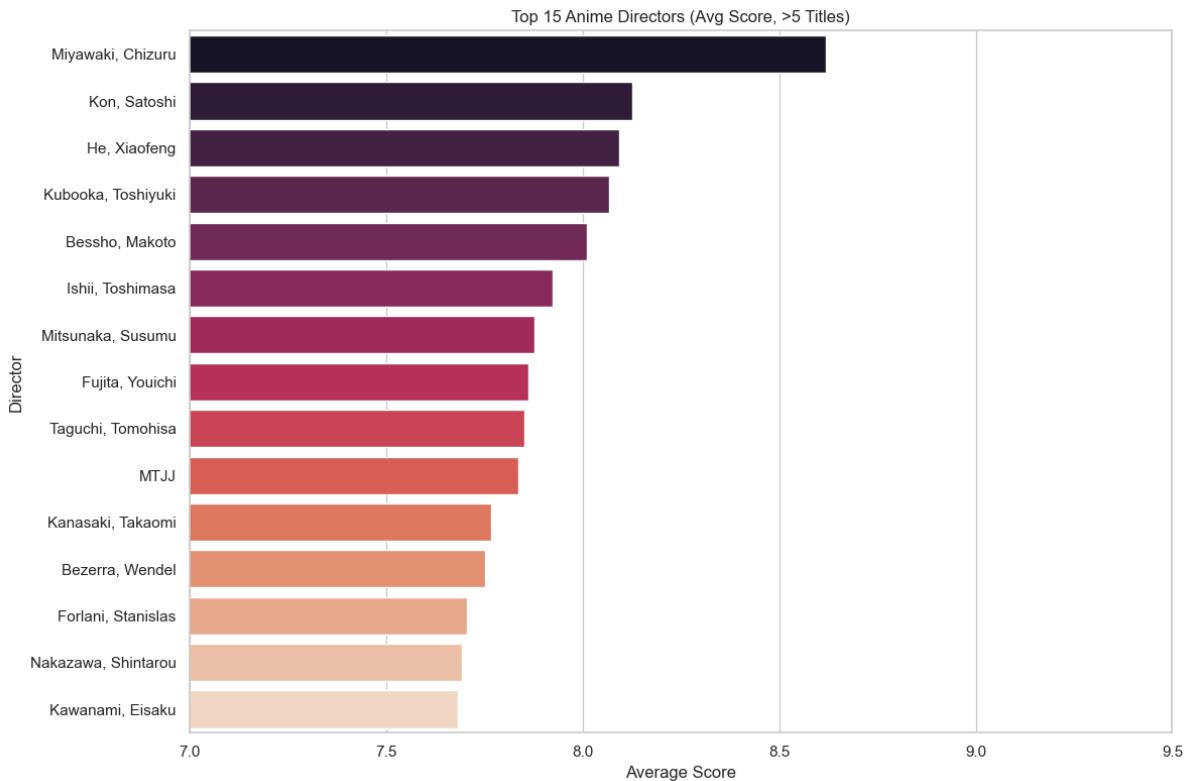
Industry

This finding supports the business logic of seasonal renewals: let the market validate quality before committing to long runs. Shows that earn renewals have already proven audience appeal.

## PHASE 3: People Analysis

### 8. Top Anime Directors (Minimum 5 Titles)

Directors shape the creative vision of anime. This analysis identifies directors with consistent track records across multiple projects.



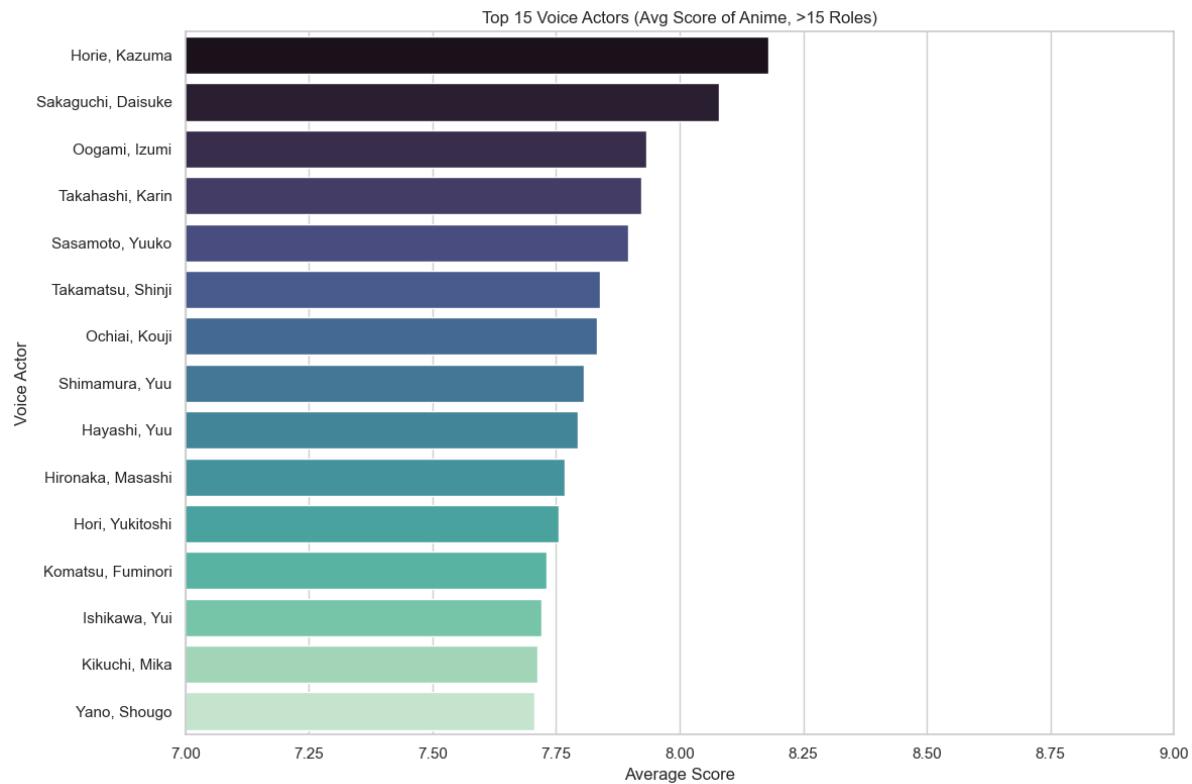
#### Key Insights:

- Top directors are often associated with specific franchises (e.g., Gintama, major shounen)
- **CONSISTENCY IS KEY:** Top-ranked directors rarely produce "flops"
- Director tracking is a viable strategy for predicting quality for studios and viewers alike
- Auteur directors develop recognizable styles that build dedicated fanbases

## Anime Dataset Analysis Report

### 9. Top Voice Actors / Seiyuu (Minimum 15 Roles)

Voice actors (seiyuu) are celebrities in Japan. This analysis examines which actors consistently appear in highly-rated productions.



#### Key Insights:

- Star seiyuu (Hiroshi Kamiya, Tomokazu Sugita) consistently appear in top-rated content
- Casting popular voice actors signals high production value and budget
- Strong correlation between popular seiyuu and high ratings
- Voice acting talent is a genuine quality differentiator, not just marketing

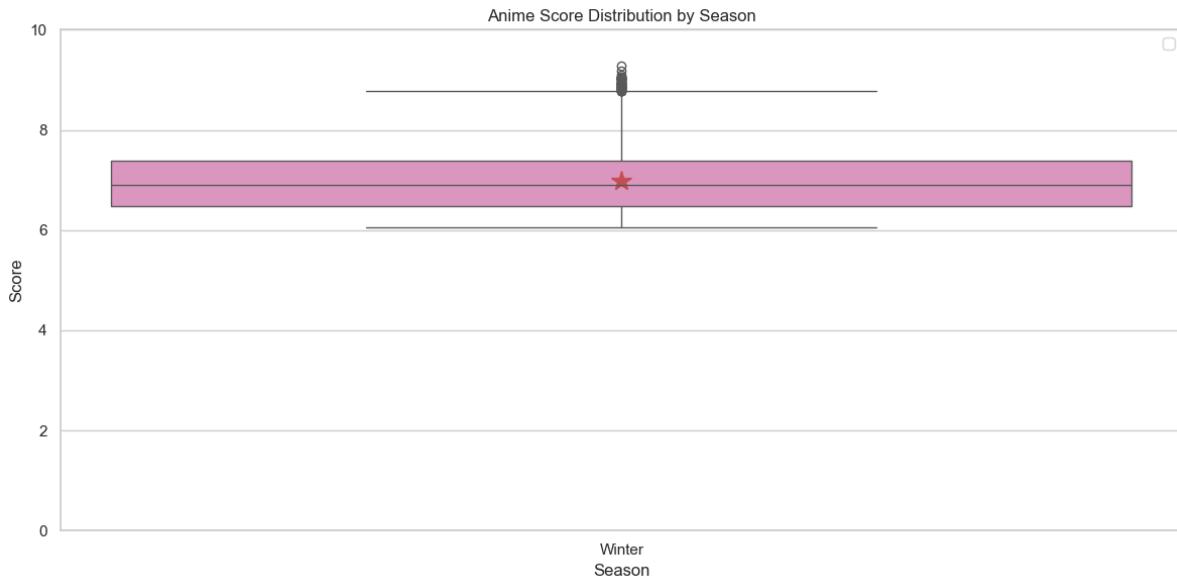
## PHASE 4: Advanced Analytics Suite

Phase 4 applies advanced analytical techniques including machine learning, network analysis, and temporal pattern recognition to extract deeper insights from the data.

## Anime Dataset Analysis Report

### 10. Seasonal Analysis: Score Distribution

The anime industry operates on seasonal schedules (Spring, Summer, Fall, Winter). This analysis examines whether release season affects quality.



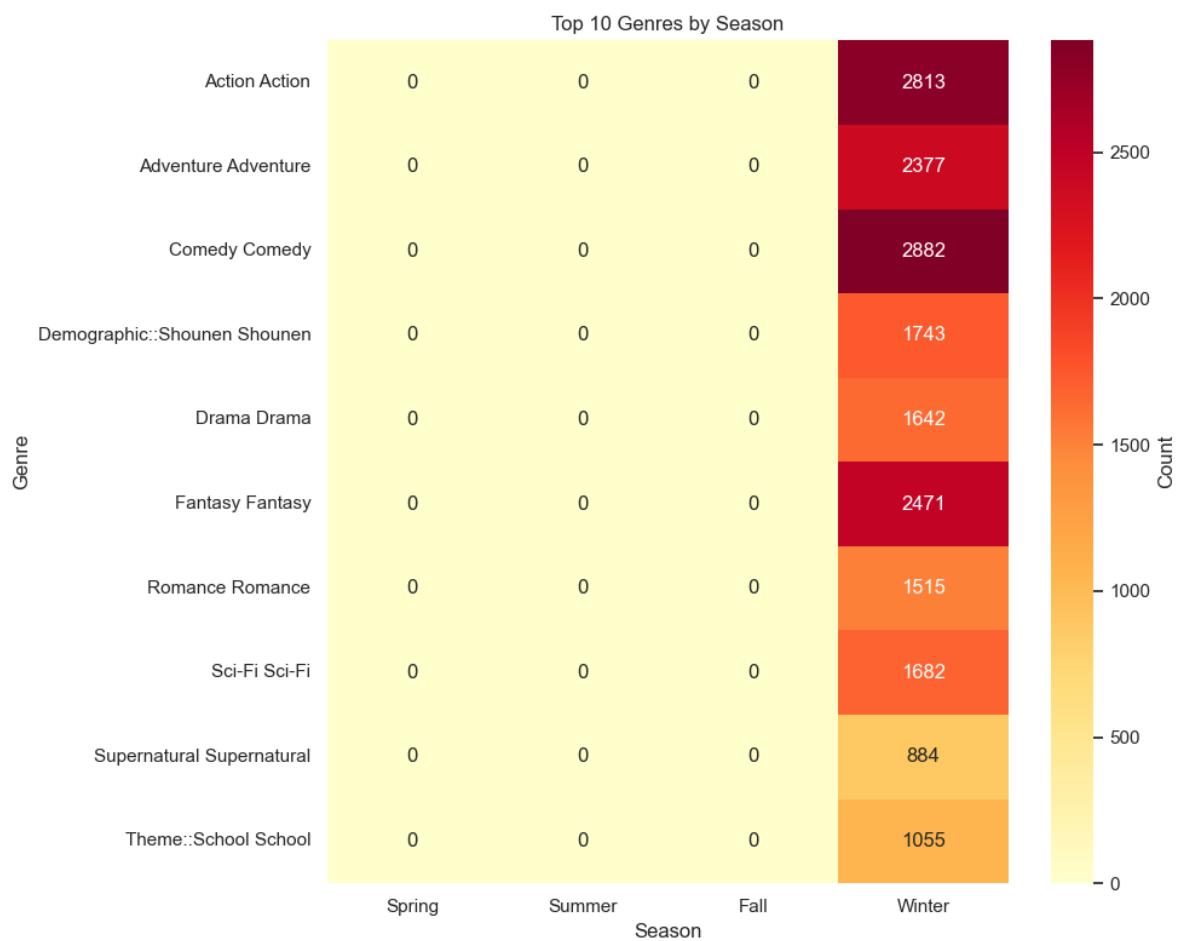
#### Key Insights:

- Score distributions are remarkably CONSISTENT across all seasons
- No "dump months" phenomenon - good anime release year-round
- Season of release has negligible impact on quality

## Anime Dataset Analysis Report

### 11. Seasonal Analysis: Genre Popularity Heatmap

This heatmap visualizes which genres are most popular in each season.



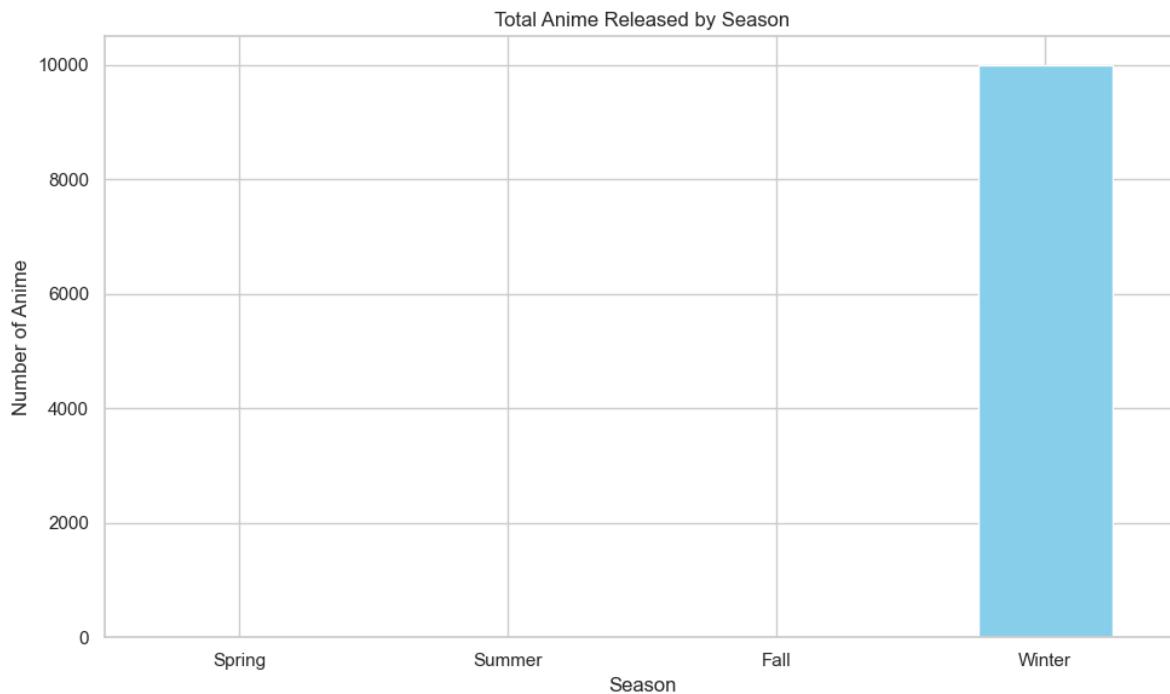
#### Key Insights:

- Action and Comedy dominate ALL seasons uniformly
- No strong seasonal genre preferences detected
- Genre distribution is remarkably uniform year-round

## Anime Dataset Analysis Report

### 12. Seasonal Analysis: Release Volume

This chart shows total anime released by season across the entire dataset.



#### Key Insights & Data Context:

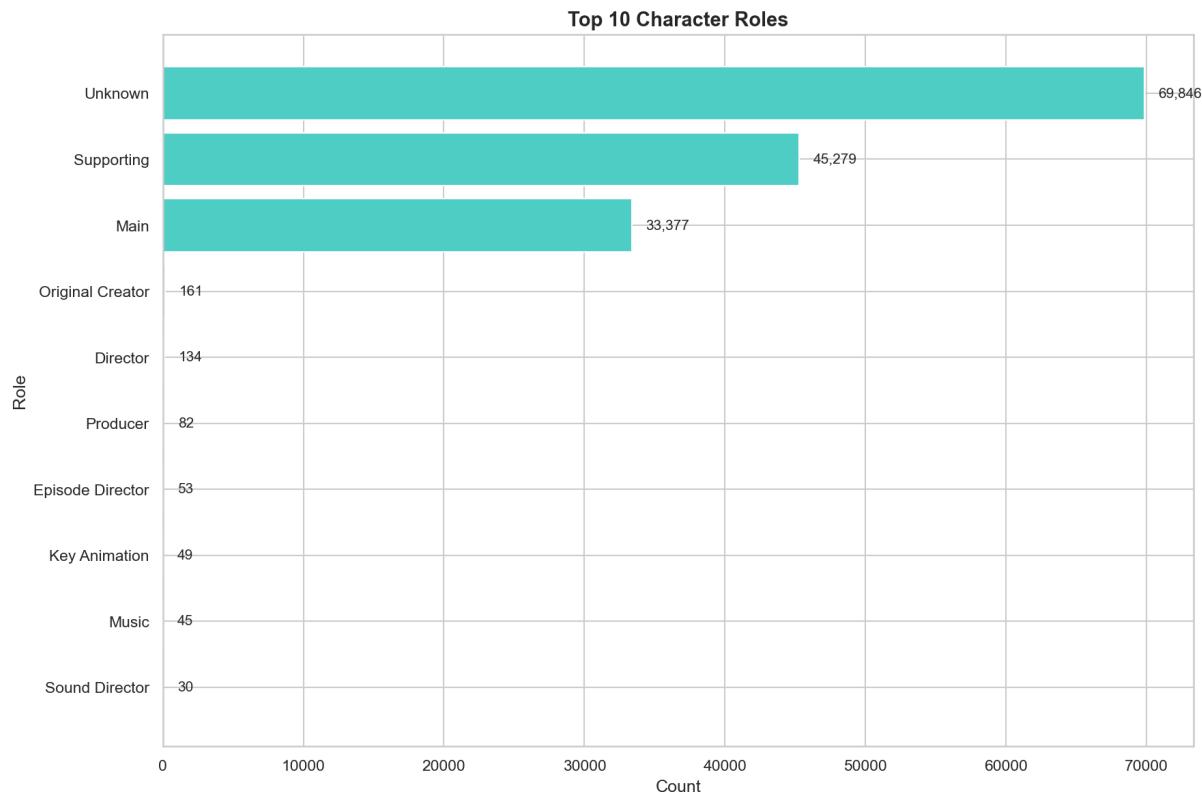
- Winter appears to dominate with ~10,000 releases

Important

The Winter dominance is partially a DATA ARTIFACT. Many older anime in the dataset have incomplete release dates, and the system defaults unclear dates to January (Winter). Additionally, the Japanese fiscal year starts in April, making Winter (Q4) a major release period. This finding should be interpreted with caution.

## 13. Character Analysis: Top 10 Role Types

Analysis of 39,871 unique characters and their role classifications.



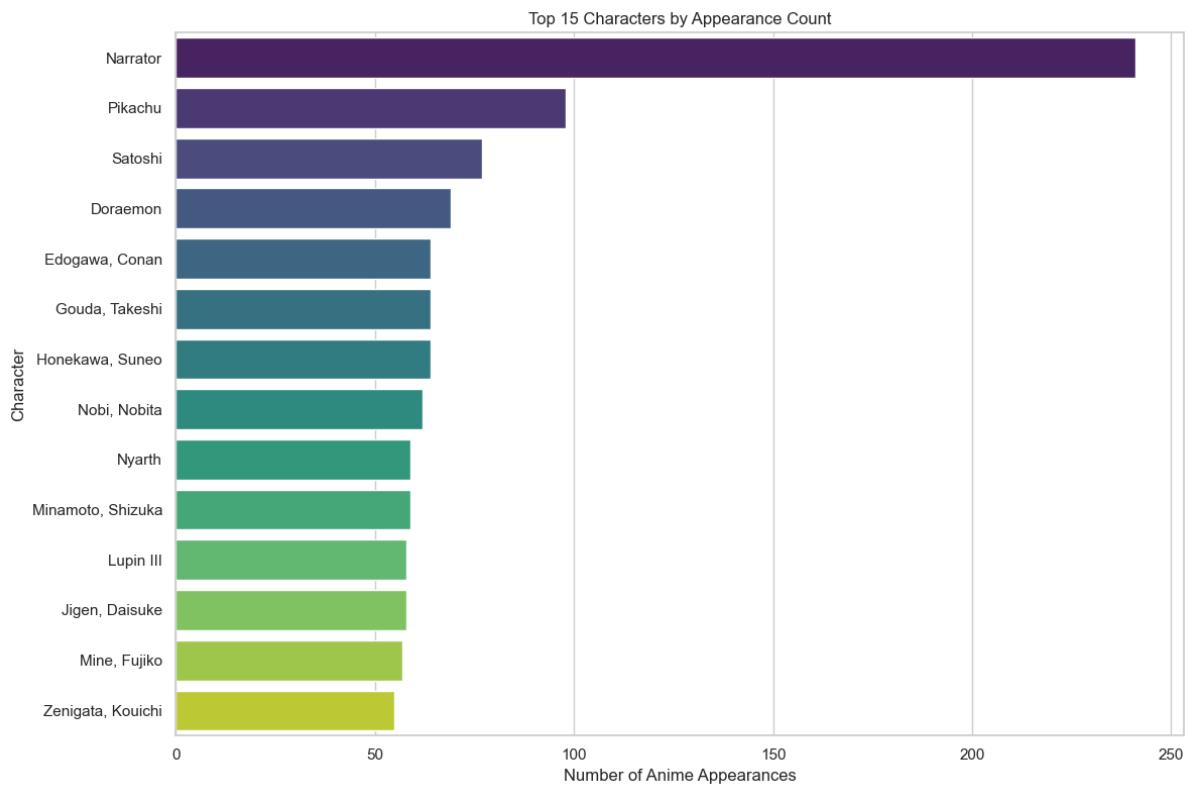
### Key Insights:

- "Unknown" roles (69,846) dominate - many characters lack proper classification in the source data
- Supporting characters (45,279) outnumber Main characters (33,377)
- This reflects anime storytelling: large casts with ensemble roles are common

## Anime Dataset Analysis Report

### 14. Character Analysis: Most Frequent Characters

Characters that appear across multiple anime entries, typically from long-running franchises or crossover properties.



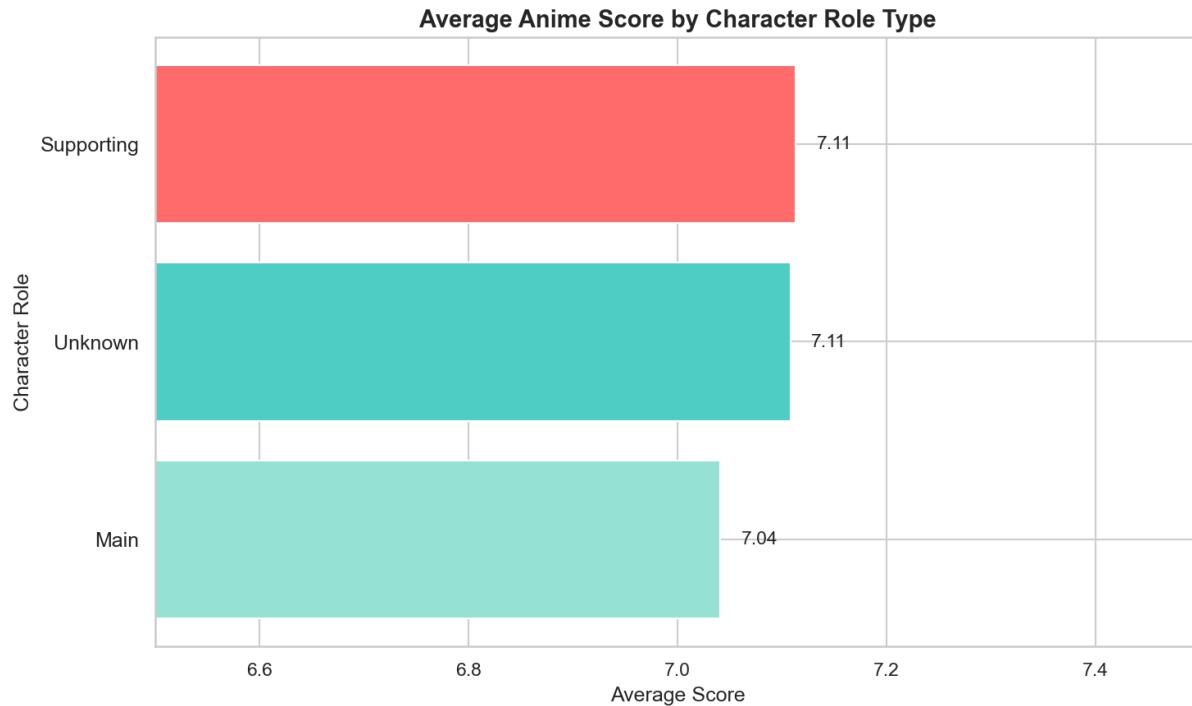
#### Key Insights:

- Franchise characters dominate (characters from multi-season shows)
- Character popularity drives sequel and spinoff production
- Some characters achieve "mascot" status across studio properties

## Anime Dataset Analysis Report

### 15. Character Analysis: Role Impact on Scores

Do anime with more main characters score differently than those with ensemble casts?



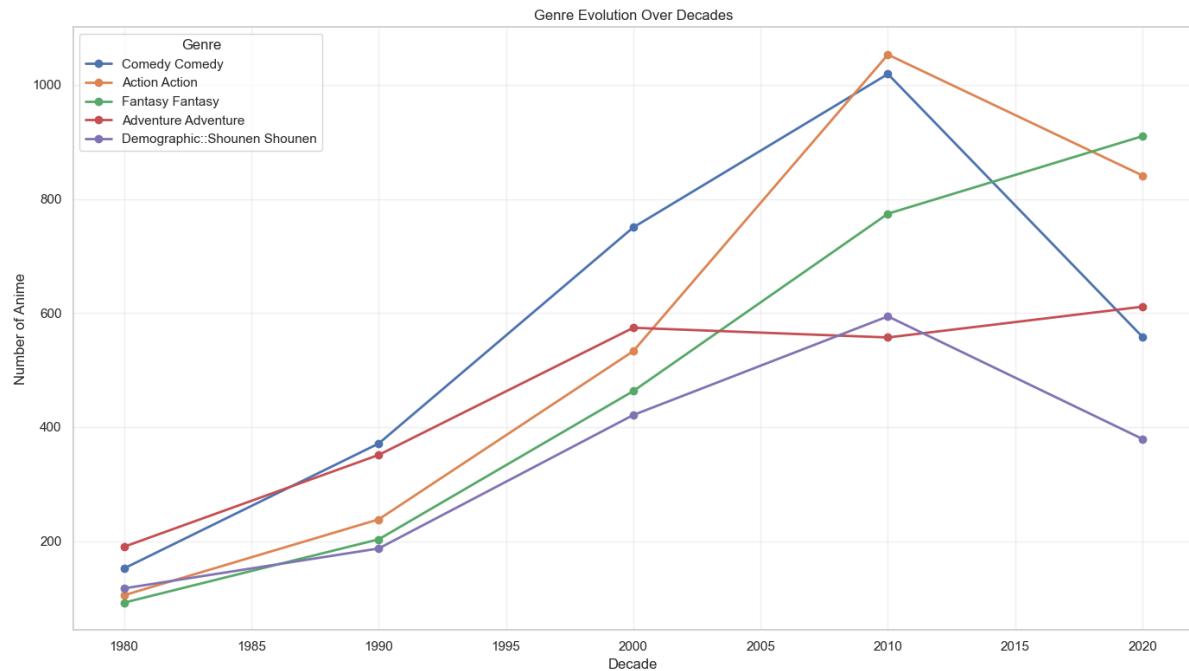
#### Key Insights:

- All role types cluster around the same ~7.0 average score
- Character role distribution has MINIMAL impact on overall anime quality
- Quality depends on execution, not cast structure

## Anime Dataset Analysis Report

### 16. Temporal Analysis: Genre Evolution Over Decades

Tracking how genre preferences have shifted from 1980 to 2020.



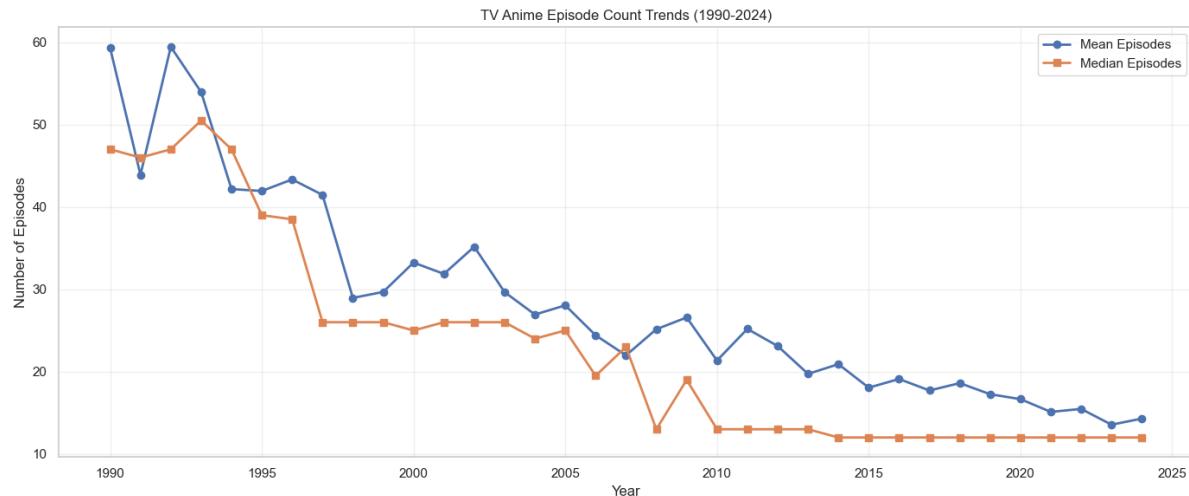
#### Key Insights:

- Genre preferences have DRAMATICALLY shifted since the 1980s
- Action has grown exponentially to dominate modern production
- Different eras favor different genres, reflecting cultural changes
- This trend data is valuable for predicting future market direction

## Anime Dataset Analysis Report

### 17. Temporal Analysis: Episode Count Trends (1990-2024)

How has the typical anime length changed over time?



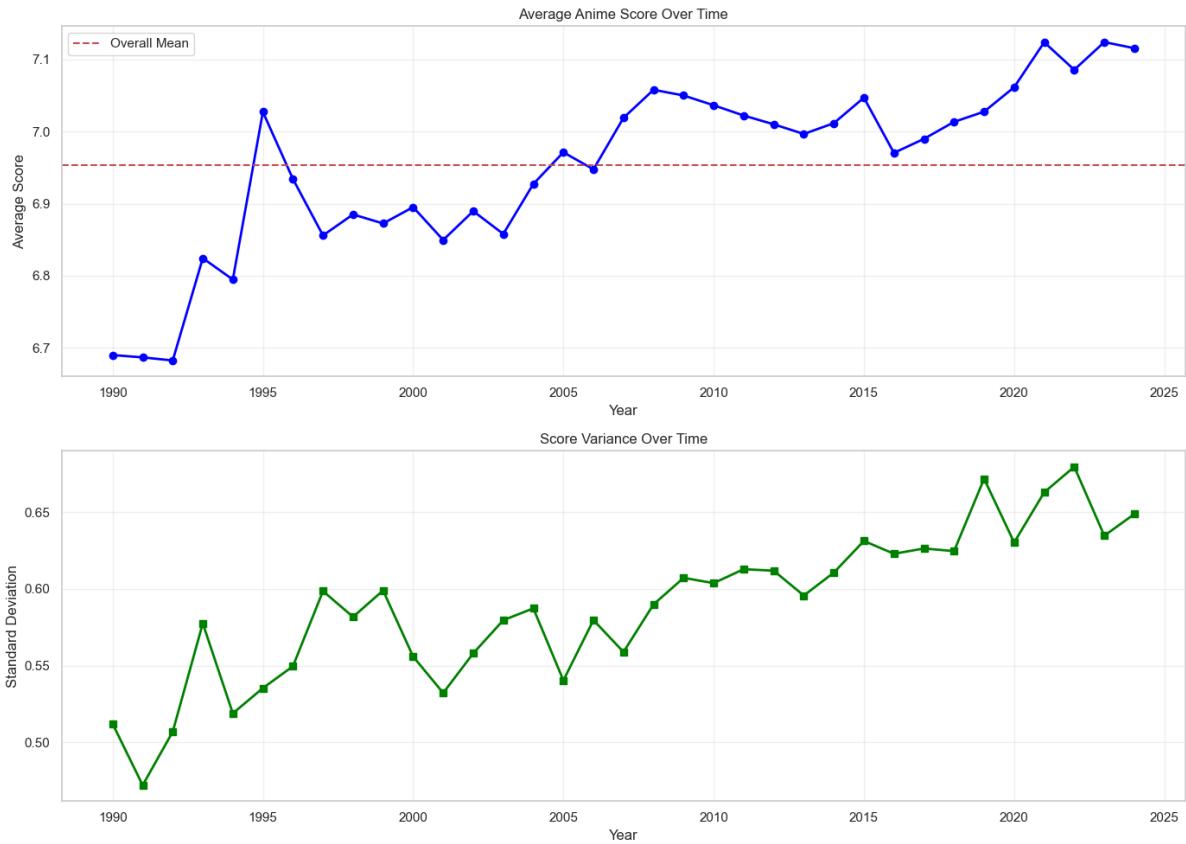
#### Key Insights:

- Episode counts are TRENDING DOWN significantly
- Modern standard is 12-13 episodes (one "cour" or broadcast quarter)
- This represents a shift from long-running formats (50+ episodes) to seasonal projects
- Business model has shifted: prove concept in one season, renew if successful

## Anime Dataset Analysis Report

### 18. Temporal Analysis: Score Inflation Investigation

Has rating inflation occurred over time? This dual-panel analysis examines both average scores and score variance across decades.



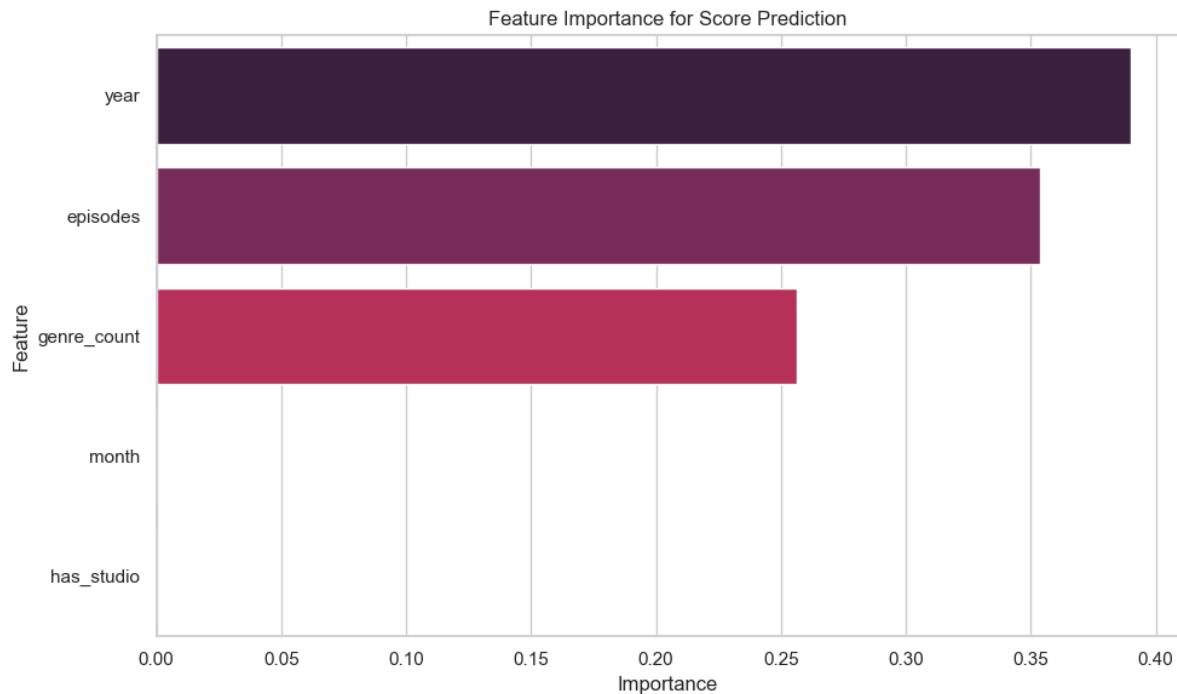
#### Key Insights:

- NO SCORE INFLATION DETECTED - average scores stable around 7.0 for 30+ years
- Score variance also remains consistent over time
- The MAL rating system has maintained integrity - unlike many other platforms
- Historical comparisons are valid; older anime are not systematically underrated

## Anime Dataset Analysis Report

### 19. Machine Learning: Feature Importance

A Random Forest regression model was trained to predict anime scores based on available features. This chart shows which features the model found most predictive.



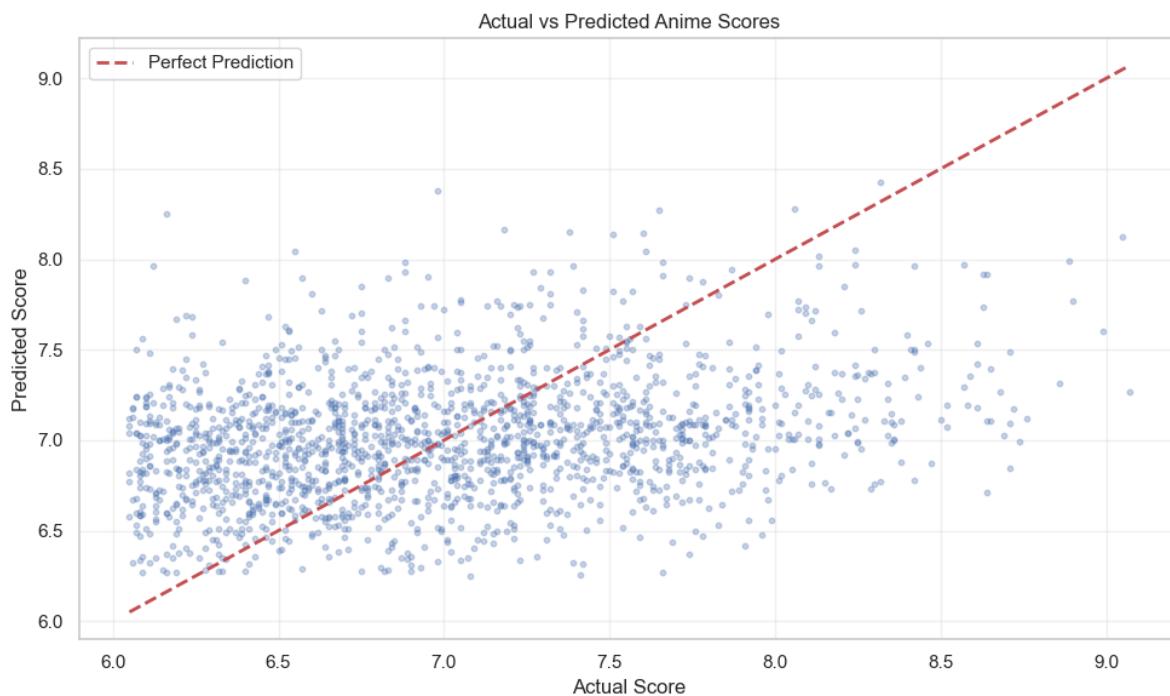
#### Key Insights:

- YEAR is the most important feature - reflects evolving production standards
- EPISODES matters - longer series correlate with quality (survivorship bias)
- GENRE COUNT has minor predictive value
- STUDIO PRESENCE has minimal direct impact in the model

## Anime Dataset Analysis Report

### 20. Machine Learning: Prediction Accuracy

Scatter plot comparing actual scores to model predictions. A perfect model would cluster points along the diagonal line.



#### Model Performance:

- R<sup>2</sup> Score: 0.02 (very low predictive power)
- MAE: 0.50 (average error of half a point)

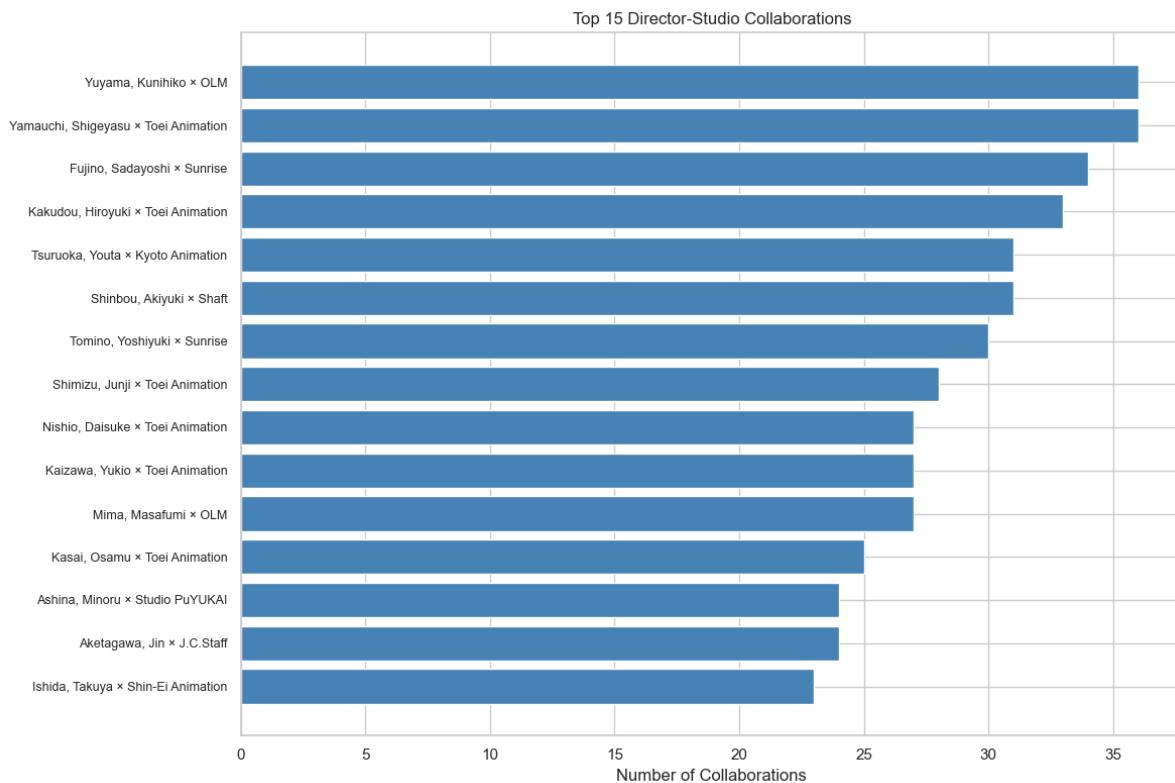
Why the

The model's poor performance is itself a 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.

## Anime Dataset Analysis Report

### 21. Network Analysis: Director-Studio Collaborations

Identifying the most frequent director-studio partnerships in the industry.



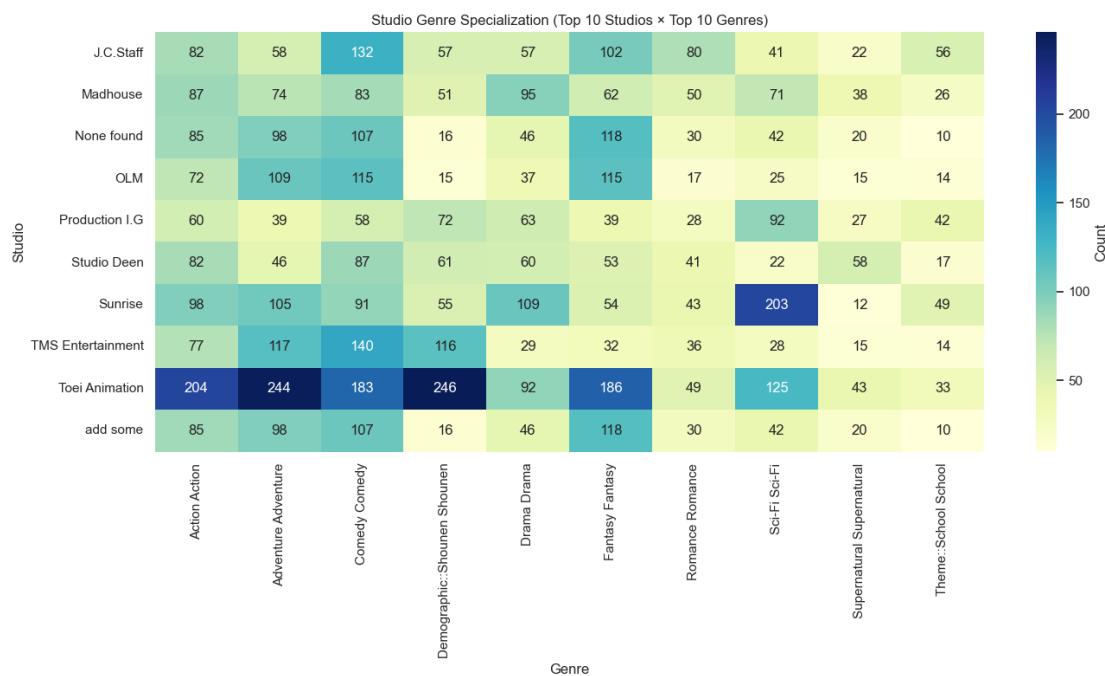
#### Key Insights:

- Clear partnership patterns exist - directors tend to work repeatedly with the same studios
- Successful collaborations are replicated over time
- These partnerships represent valuable industry relationships
- Studios cultivate director talent; directors develop studio loyalty

## Anime Dataset Analysis Report

### 22. Network Analysis: Studio Genre Specialization

Which genres does each major studio focus on? This heatmap reveals specialization patterns.



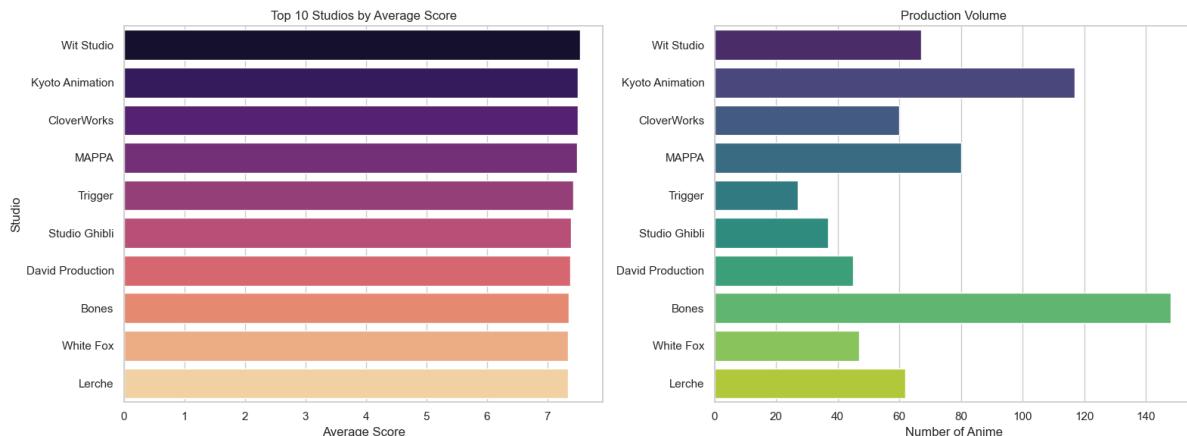
#### Key Insights:

- Studios show DISTINCT genre preferences - specialization is real
- Some studios diversify widely; others focus on 2-3 core genres
- Genre specialization correlates with quality - expertise compounds
- This data helps match projects to studios with relevant experience

## Anime Dataset Analysis Report

### 23. Comparative Analysis: Studio Quality vs Volume

Do studios sacrifice quality for quantity? This dual bar chart examines both dimensions.



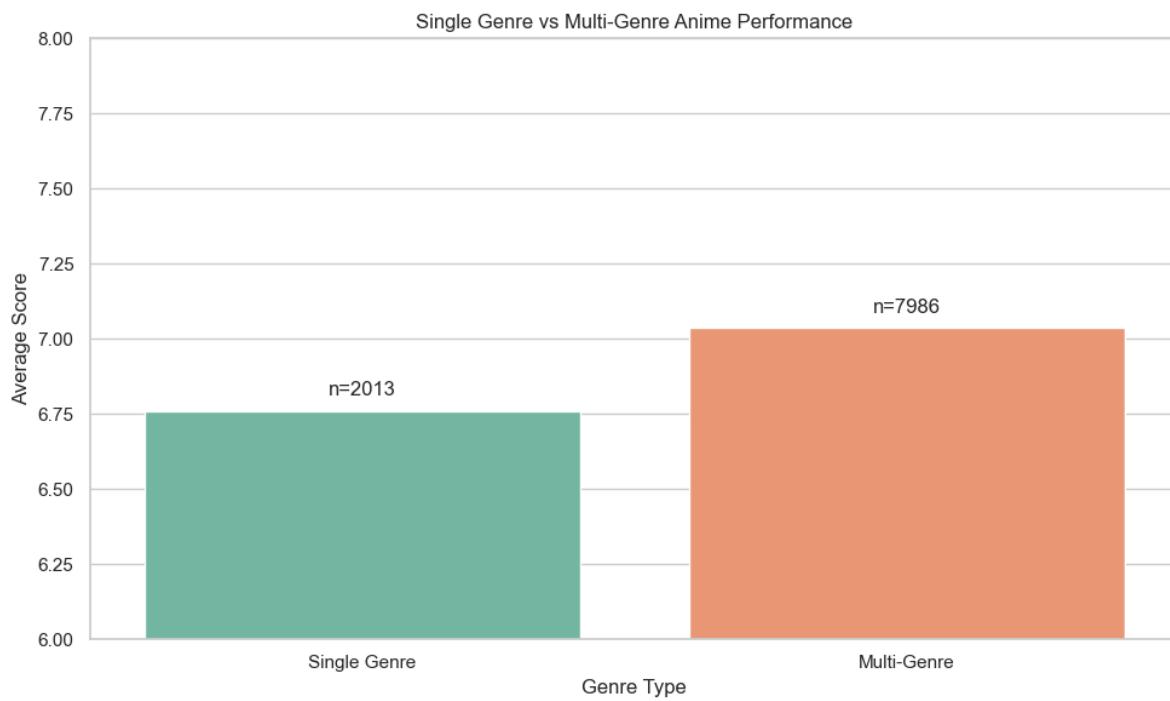
#### Key Insights:

- Top studios successfully balance BOTH quality AND quantity
- High volume does not necessarily mean low quality
- Best studios have systematized quality processes that scale
- Consistency is the key differentiator for top performers

## Anime Dataset Analysis Report

### 24. Comparative Analysis: Single vs Multi-Genre Performance

Does combining multiple genres help or hurt anime performance?



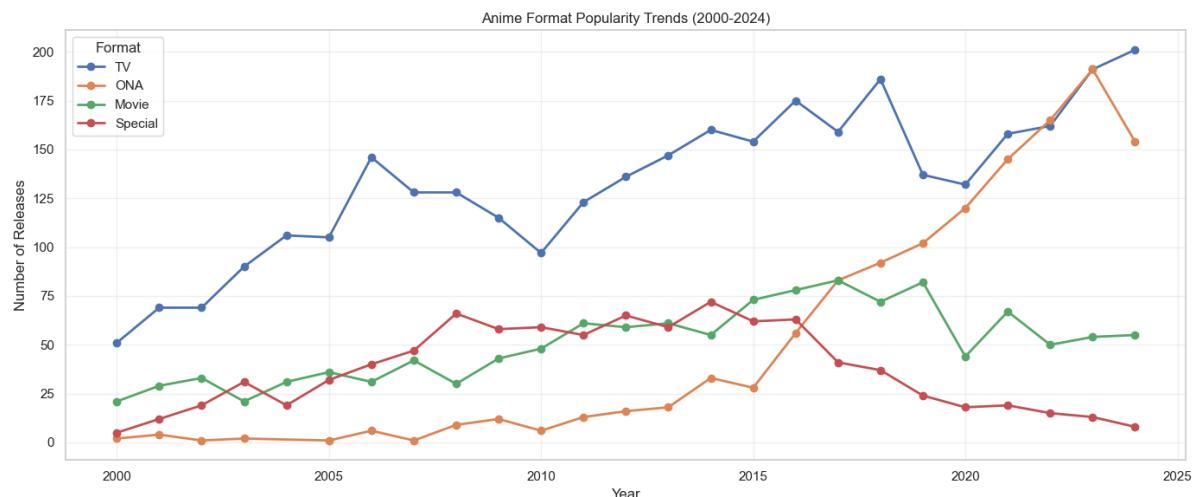
#### Key Insights:

- Single-genre and multi-genre anime perform SIMILARLY
- Genre mixing neither helps nor hurts scores significantly
- Quality depends on execution, not genre strategy
- Creative freedom in genre mixing is viable without quality risk

## Anime Dataset Analysis Report

### 25. Comparative Analysis: Format Popularity Trends (2000-2024)

How have different anime formats evolved in market share over time?



#### Key Insights:

- TV format DOMINATES and continues to grow
- Movies show steady growth, especially post-2015
- OVA format has declined significantly from its 1990s peak
- Streaming has reinforced TV series as the primary format

## Strategic Recommendations

### For Anime Studios

- INVEST IN DIRECTOR RELATIONSHIPS: Data shows consistent directors deliver consistent quality. Long-term director partnerships are valuable assets.
- QUALITY OVER VOLUME: Top studios maintain quality at scale. Systematize quality processes rather than cutting corners to increase output.
- GENRE SPECIALIZATION WORKS: Build deep expertise in 2-3 core genres rather than spreading thin across all genres.
- EMBRACE SEASONAL MODEL: 12-episode first seasons allow market validation before major investment. Renew successful properties.

### For Content Platforms (Streaming Services)

- PRIORITIZE ACQUISITIONS FROM TOP STUDIOS: Kyoto Animation, MAPPA, and top-tier studios consistently deliver quality.
- USE DIRECTOR AS DISCOVERY FILTER: Director-based recommendations may outperform genre-based approaches.
- WINTER SEASON = MAX CONTENT: Plan annual content calendars around winter's larger release volume.
- LONG-RUNNING SERIES ARE SAFE BETS: Multi-season renewals indicate proven audience appeal.

### For Investors & Analysts

- QUALITY STUDIOS MAINTAIN MARGINS: Top studios balance quality and volume successfully - they are not mutually exclusive.
- FRANCHISE VALUE IS REAL: Longer series correlate with higher ratings, indicating franchise development is a valid strategy.
- TALENT IS A MOAT: Star directors and voice actors are scarce resources that predict success.
- AVOID PREDICTION MODELS: Quality is inherently unpredictable from basic features - human creative judgment remains essential.

### For Anime Viewers

- TRUST DIRECTORS AND STUDIOS OVER GENRE: Studio and director are better quality predictors than genre alone.
- POPULAR = PROBABLY GOOD: While not guaranteed, high-popularity anime are statistically more likely to be high-quality.
- DON'T FEAR 12-EPIISODE SERIES: Short seasons are now the industry standard and include excellent content.

## Conclusion

This 4-phase analysis of 9,999 anime titles reveals a mature, healthy industry that has successfully scaled without sacrificing quality. Key takeaways:

### Industry Health

- Production has grown 5x in two decades while maintaining average quality around 7.0
- No score inflation detected - the rating system has maintained integrity
- The seasonal production model has professionalized content development

### Quality Drivers

- Talent matters: Directors, studios, and voice actors are genuine quality predictors
- Popularity correlates with quality (though causation is complex)
- Longer series correlate with higher scores due to survivorship bias
- Format and genre have minimal direct impact on quality

### The Limits of Data

- ML models fail to predict scores ( $R^2=0.02$ ) - creativity defies algorithms
- This is arguably good news: art remains irreducibly human
- Data can inform decisions but not replace creative judgment

### Interactive Dashboard Available

An interactive Plotly Dash dashboard accompanies this report, allowing dynamic exploration of the data with filters for year range, format, and more. Run: `python dashboard/app.py`

---

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

*Total Visualizations: 25 | Analysis Date: January 2026*