Multivariate Analysis Report on Anime Dataset

ANKIT KUMAR (23114006)

March 2025

Contents

1 Introduction

Anime has grown into a global phenomenon, captivating audiences across various demographics with its diverse storytelling, unique art styles, and engaging character development. With the ever-expanding anime industry, understanding trends, audience preferences, and performance indicators has become crucial for producers, streaming platforms, and fans alike.

This report presents a **multivariate analysis** of a comprehensive anime dataset, encompassing various attributes such as genres, ratings, popularity, types, and production studios. By exploring **univariate, bivariate, and multivariate relationships**, we aim to uncover patterns and insights that drive the success and reception of anime titles.

Objectives of the Study:

- To analyze the distribution and popularity of different anime genres.

- To identify the relationship between **score, popularity, and member engagement**.

- To explore how different **anime types (TV, OVA, movie)** perform in terms of user ratings and votes.

- To assess the influence of **studios and source material** on the success of an anime.

- To provide actionable insights that can guide industry stakeholders in decision-making.

By leveraging statistical analysis and visual representations, this report provides a detailed examination of how various factors interplay in determining the success of anime titles. The findings may be valuable for anime studios, streaming services, and fans seeking data-driven insights into the industry.

2 Dataset Description

The dataset consists of 24,030 anime entries with 16 attributes, providing information about various anime series, movies, and specials. Below is a summary of the dataset:

Column	Data Type	Description
id	Integer	Unique identifier for each anime.
Title	String	Name of the anime.
Votes	Float	Number of votes received for scoring.
Status	String	Airing status (e.g., finished airing, currently airing).
Types	String	Type of anime (e.g., TV, movie, OVA, special).
Episodes	Float	Number of episodes in the anime.
Score	Float	Average user rating score.
Source	String	Original source material (e.g., manga, novel, game).
Members	Float	Number of users who have added this anime to their lists.
Popularity	Float	Popularity ranking based on member count.
Favourites	Float	Number of users who have marked this anime as a favorite.
Rating	String	Age rating of the anime (e.g., PG, R, R+).
Year	Float	Year in which the anime was released.
Season	String	Season of release (e.g., spring, summer, fall, winter).
Genres	String	List of genres associated with the anime.
Studios	String	List of animation studios involved in production.

3 Dataset Statistics

- **Total Entries:** 24,030

- **Missing Data:** Some fields like "Episodes", "Score", and "Studios" contain missing values.

- **Score Range:** 1.85 - 9.14 (Mean: 6.46)

- **Popularity Range:** 1 - 20,538

- **Year Range:** 1917 - 2023

4 Univariate Analysis

4.1 Top Genres Distribution (Bar Chart)

- **Comedy, Action, and Fantasy** are the most frequent genres.

- **Military, Mystery, and Ecchi** are among the least common genres.

- The distribution suggests that viewers prefer **light-hearted and action-packed content** over niche genres.

4.2 Anime Status (Bar Chart)

- The vast majority of anime are *Finished Airing* (**99.5%**).

- A very small percentage is currently airing (**0.4%**) or not yet aired (**0.1%**).

- This indicates that most anime data available focuses on **completed series rather than ongoing or upcoming ones**.

4.3 Anime Type Distribution (Bar Chart)

- **OVA (Original Video Animation)** is the most common type, followed by TV series.

- Movies, ONAs (Original Net Animation), and Specials are also common.

- The dominance of OVAs suggests a strong market for **direct-to-video content**, alongside TV series.

5 Bivariate Analysis

5.1 Correlation Matrix Analysis

- **Votes and Members:** A very high positive correlation (**0.99**) indicates that anime with more members tend to receive more votes.

- **Favourites and Members:** A strong correlation (**0.79**) suggests that highly followed anime are often favorited.

- **Popularity and Score:** A moderate negative correlation (**-0.51**) suggests that highly popular anime do not always receive the best ratings.

5.2 Scatter Plots (Score vs. Members, Score vs. Votes)

- A positive trend exists between **Score and Members**, meaning higher-rated anime tend to attract more members.

- The relationship is **weakly positive**, indicating that score alone does not determine membership size.

- Votes also increase with Score, but there are many outliers where low-scoring anime still received significant votes.

5.3 Score Distribution by Rating Category (Violin & Boxplots)

- **R-rated anime** tend to have the highest score variability.

- **G-rated and PG-rated anime** generally have more consistent average scores, indicating they cater to a broader, family-friendly audience.

5.4 Votes by Anime Type (Boxplot)

- **TV series** get the most votes, followed by **movies and OVAs**.

- **Music-based anime and web animations (ONA)** tend to have lower votes, likely due to niche audiences.

6 Multivariate Analysis

6.1 Votes by Anime Source (Violin Plot)

- **Manga adaptations** get the highest average votes, meaning manga-based anime tend to be more popular.

- **Original anime** have a wider range of votes, suggesting unpredictability in audience reception.

6.2 Top Studios by Average Votes

- **Madhouse, Square Enix, and Studio 4°C** rank among the top studios based on average votes.

- High variability in votes suggests that even top studios produce both highly successful and lower-rated projects.

7 Conclusions & Recommendations

- **Popular genres** (Comedy, Action) can be targeted for content recommendations or streaming platform marketing.

- Anime studios may focus on **Fall releases**, as it appears to be the most active season.

- **TV series and OVAs dominate** the industry, indicating strong audience demand for both traditional and digital-first releases.

- Long-running series tend to have more **variability in ratings**, suggesting that maintaining quality over many episodes is challenging.

- Highly popular anime (with many members) do not necessarily get the highest ratings, meaning studios should **balance audience reach with content quality**.

- **Manga adaptations perform best**, so new anime projects based on popular manga may have higher chances of success.

- **Certain studios consistently receive high votes**, but variability exists, meaning detailed audience analysis is needed for better market success.