

## 1. Dataset Description

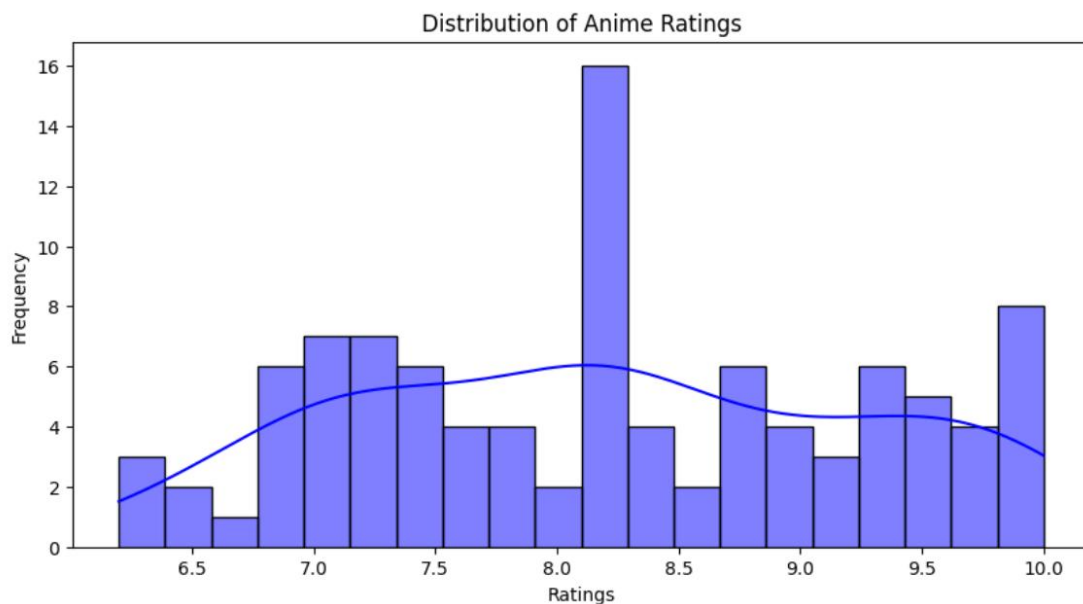
We selected the dataset "Most Watched Anime Dataset", This dataset contains 100 records and 9 features, covering the most popular anime names, countries, ratings, number of episodes, animation studios, budget, release years, genres, and episode durations.

During data cleaning, we found that some columns (such as ratings, animation studios, and budget) had missing values. Different filling strategies were applied, such as using the median for numerical data and filling unknown categories with "Unknown", and converting data types to fit visualization needs.

## 2. Visualization Display

We created the following four different visualization charts:

### 2.1 Rating Distribution (Histogram)



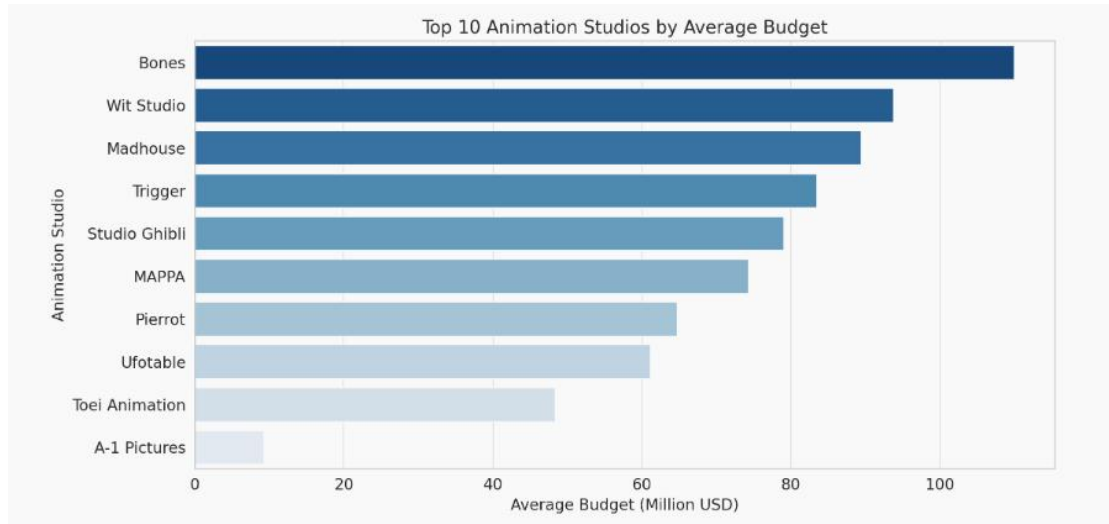
Objective: To observe the overall distribution of anime ratings.

Analysis:

- Ratings are mainly concentrated between 6.0 - 9.5.
- A few anime have lower ratings (<6.5), but the quantity is small.

- The KDE density curve shows the rating distribution trend.

## 2.2 Comparison of Budgets Among Different Animation Studios (Bar Chart)

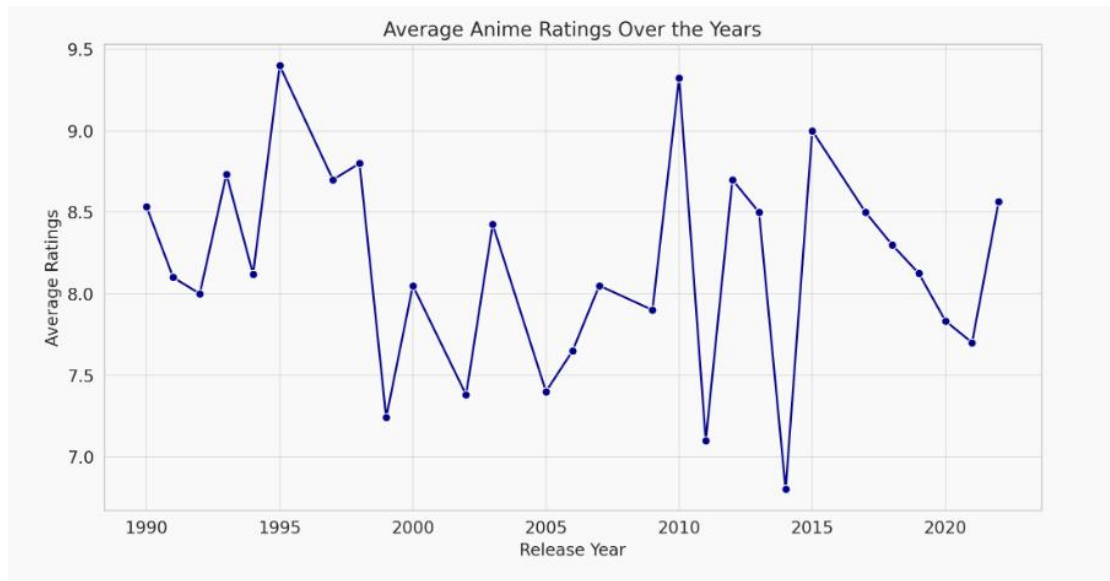


Objective: To compare the budget levels of different animation studios.

Analysis:

- The animation studios with the highest budgets are in the range of 80-100 million USD.
- There is a large budget difference between studios, which may affect the quality of animation production.

## 2.3 Release Year vs. Rating (Line Chart)

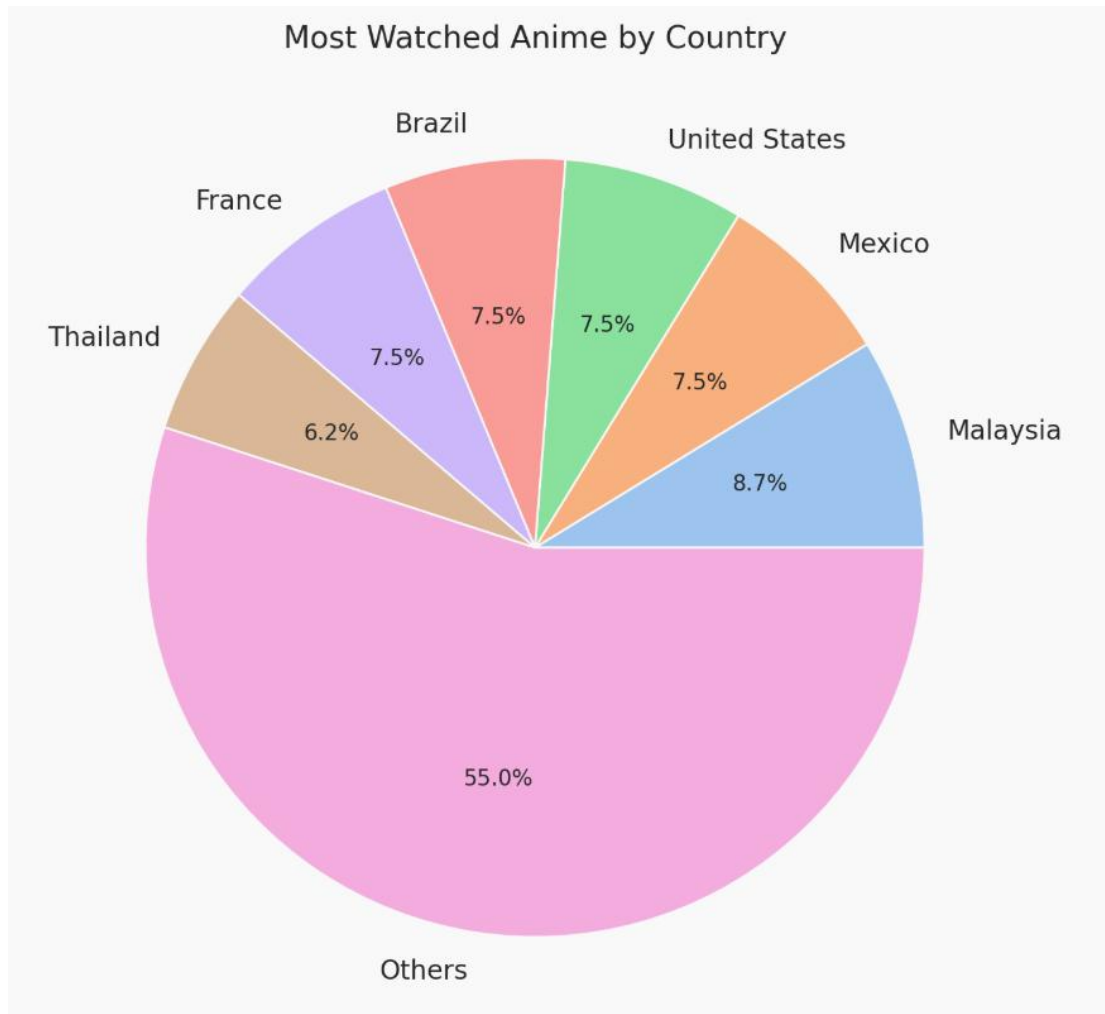


Objective: To observe the trend of anime ratings over time.

Analysis:

- Ratings fluctuate in different years.
- Some years have overall higher ratings, possibly due to a larger number of high-quality anime productions.

## 2.4 Distribution of the Most Popular Anime by Country (Pie Chart)



Objective: To show the proportion of most-watched anime in different countries.

Analysis:

- A few countries dominate the viewership, reflecting their enthusiasm for anime.
- Viewership in other countries is more dispersed.

### 3. Methodology

We used Python for data cleaning and visualization. The specific methods are as follows:

#### 1. Data Cleaning:

- Removing rows with excessive missing values (e.g., where Anime Name is empty).
- Using the median to fill in missing numerical values,"Unknown"filling missing category values.

- Converting numerical columns to the correct data type.

## 2. Data Visualization:

- Histogram: Using Seaborn's histplot() to plot the rating distribution.
- Bar Chart: Using barplot() to compare budgets among different animation studios.
- Line Chart: Using lineplot() to show rating trends over time.
- Pie Chart: Using Matplotlib's pie() to display the country distribution.

## 4. Libraries Used

The main Python libraries used in this project are:

- pandas for data processing and cleaning.
- matplotlib 和 seaborn for data visualization.

## 5. Results Analysis

Through visualization, we obtained the following conclusions:

- Rating distribution: Most anime have high ratings, likely because popular works usually receive high scores.
- Animation studio budgets: Studios with higher budgets may have greater production capabilities and better quality works.
- Ratings over time: Some years have higher ratings, possibly indicating peak periods in the anime industry with a large number of high-quality works.
- Country distribution: Some countries have a higher acceptance of anime, which may influence market trends.

## 6. Team Contributions

Data Collection:HaoranSun A20563431

Code Writing:GuodaYou A20563446

Report Writing:YongzhiTong A20563466