

The growing
phenomena of

ANIME WEDDING

Yeng Russian



Agenda

DATA PRESENTATION

- *Introduction to the topic:*
 - *Anime definition.*
 - *Characteristics.*
- *Importance of Research.*
- *Research Questions.*
- *Data Overview.*
- *Data pre-processing*
 - *Clean the data (NaN values, duplicates and more).*

DATA PROCESSING

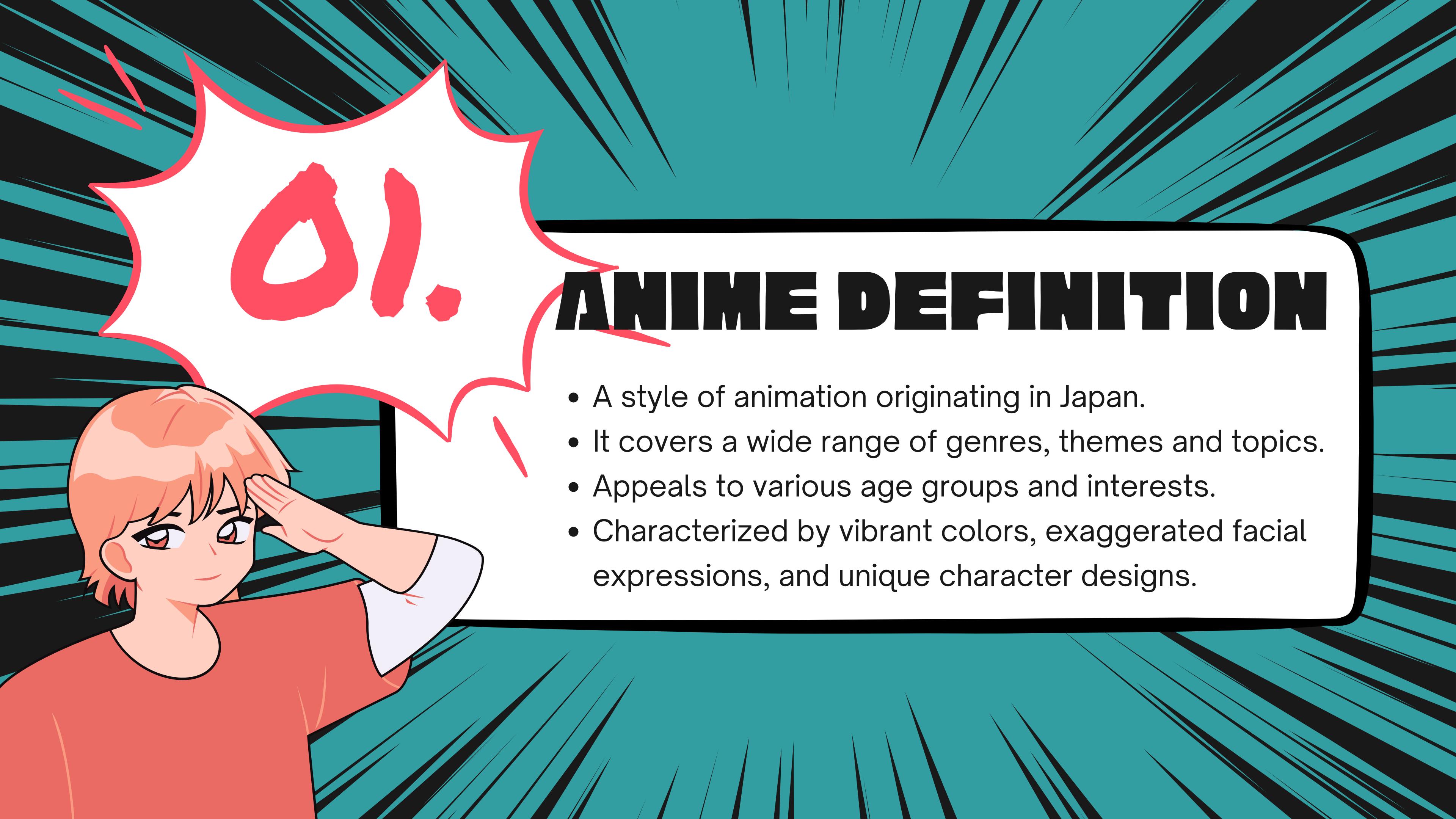
- Manipulate the data to answer the Research Questions.
- Present pertinent Graphs.
- Explain the results.



CLOSING STATEMENT

- *Conclusions.*
- *Limitations.*





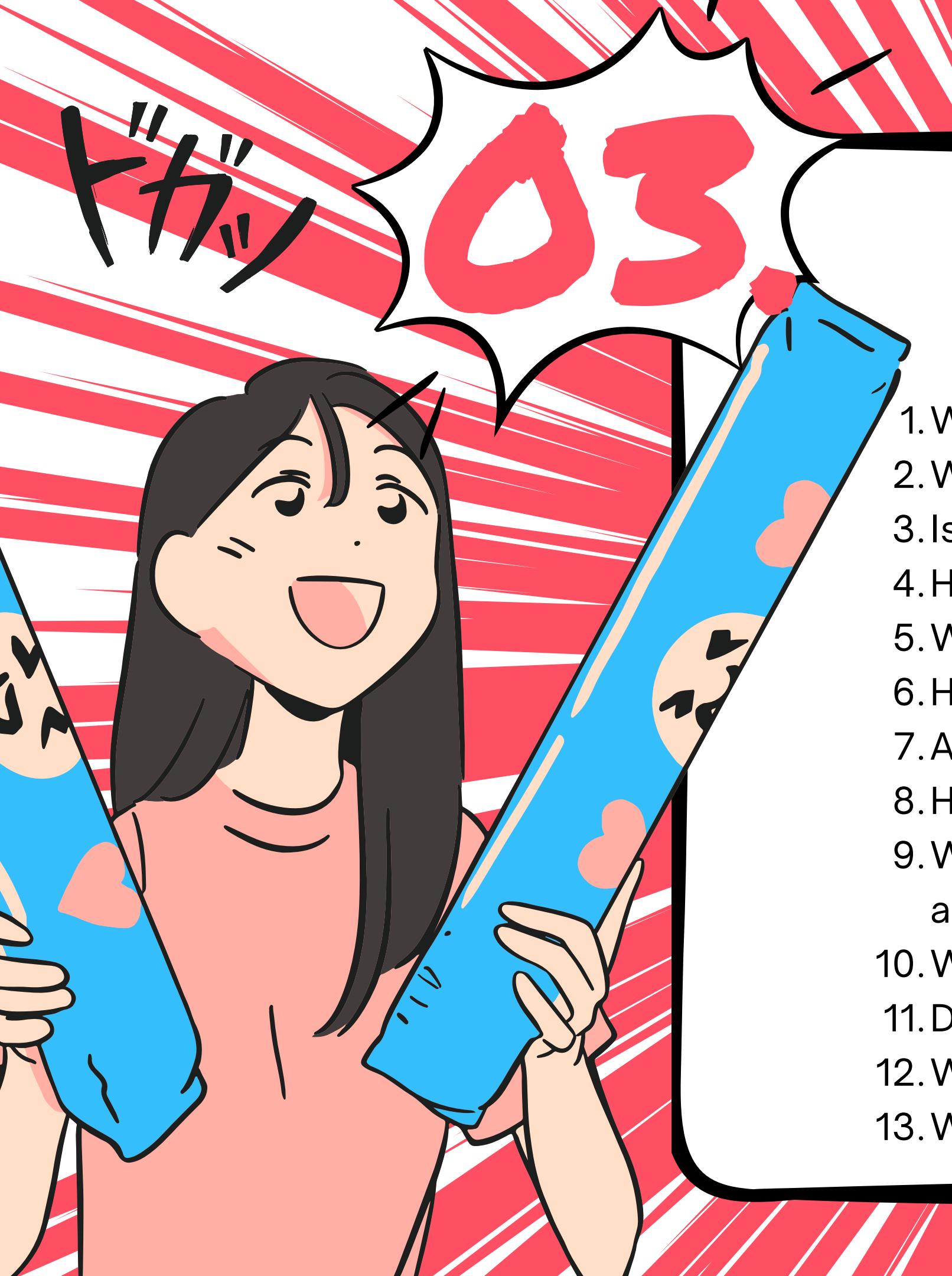
ANIME DEFINITION

- A style of animation originating in Japan.
- It covers a wide range of genres, themes and topics.
- Appeals to various age groups and interests.
- Characterized by vibrant colors, exaggerated facial expressions, and unique character designs.



Importance of Research

- Make informed decisions about the anime industry.
- Identify Audience Preferences.
- Identify Popularity Trends.
- Guide the decision-making process and generate business value.
- Investment Opportunities



Research Questions

1. Which animes have the highest and lowest user scores?
2. What is the relationship between Score and Votes?
3. Is there any anime with a high number of Votes but a low Score?
4. How does Popularity rank correlate with Score?
5. What are the most common age Rating categories?
6. How has the average score of anime evolved?
7. Are animes with more Episodes generally higher scored?
8. How does the status of the anime affect the score?
9. What are the most common Source materials for high-scoring anime?
10. What are the most common duration of Episodes?
11. Do shorter or longer anime (Duration) tend to have higher scores?
12. What are the most common episode counts?
13. What is the relationship between Popularity and Votes?



DATA OVERVIEW

- Title: Title of the anime (Jap & Eng).
- Score: Score given by users (out of 10).
- Votes: Number of user votes for the anime.
- Ranked: Rank of the anime based on score.
- Popularity: Popularity rank.
- Episodes: Number of episodes.
- Status: Current airing status (e.g., Finished Airing).
- Aired: Airing period.
- Premiered: Premiere season and year.
- Producers: Production companies.
- Licensors: Licensing companies.
- Studios: Animation studios.
- Source: Source material (e.g., Manga).
- Duration: Duration per episode.
- Rating: Age rating.

Read Dataset

Quick overview

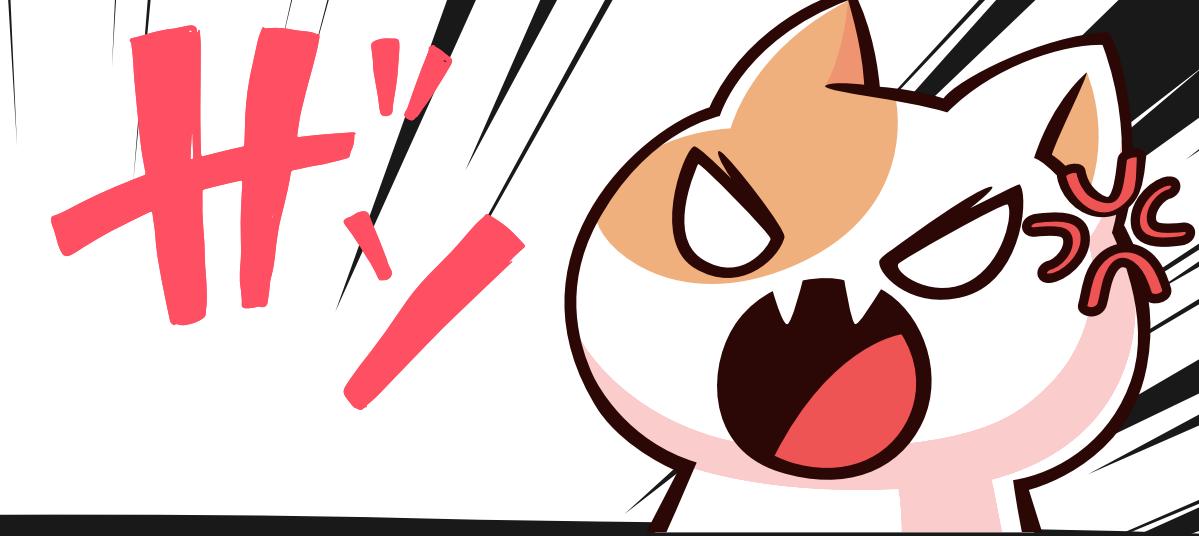
```

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Title        10000 non-null   object  
 1   Score        10000 non-null   float64
 2   Vote         10000 non-null   int64   
 3   Ranked       10000 non-null   int64   
 4   Popularity   10000 non-null   int64   
 5   Episodes     10000 non-null   object  
 6   Status       10000 non-null   object  
 7   Aired        10000 non-null   object  
 8   Premiered    4878 non-null    object  
 9   Producers    10000 non-null   object  
 10  Licensors    10000 non-null   object  
 11  Studios      10000 non-null   object  
 12  Source       10000 non-null   object  
 13  Duration     10000 non-null   object  
 14  Rating       9980 non-null    object  
dtypes: float64(1), int64(3), object(11)
memory usage: 1.1+ MB

```

	Title	Score	Vote	Ranked	Popularity	Episodes	Status	Aired	Premiered	Producers	Licensors	Studios	Source	Duration	Rating
0	Sousou no Frieren Frieren: Beyond Journey's End	9.14	128768	1	508	28	Currently Airing	Sep 29, 2023 to Mar 24, 2024	Fall 2023	'Aniplex', 'Dentsu', 'Shogakukan-Shueisha Pro... None found, add some	Madhouse	Manga	24 min. per ep.	PG-13 - Teens 13 or older	
1	Fullmetal Alchemist: Brotherhood	9.09	2080863	2	3	64	Finished Airing	Apr 5, 2009 to Jul 4, 2010	Spring 2009	'Aniplex', 'Square Enix', 'Mainichi Broadcast...' Funimation, Aniplex of America	Bones	Manga	24 min. per ep.	R - 17+ (violence & profanity)	
2	Steins;Gate	9.07	1375512	3	13	24	Finished Airing	Apr 6, 2011 to Sep 14, 2011	Spring 2011	[Frontier Works', 'Media Factory', 'Kadokawa ...] Funimation	White Fox	Visual novel	24 min. per ep.	PG-13 - Teens 13 or older	
3	Gintama®Gintama Season 4	9.06	246431	4	337	51	Finished Airing	Apr 8, 2015 to Mar 30, 2016	Spring 2015	[TV Tokyo', 'Aniplex', 'Dentsu'] Funimation, Crunchyroll	Bandai Namco Pictures	Manga	24 min. per ep.	PG-13 - Teens 13 or older	
4	Shingeki no Kyojin Season 3 Part 2Attack on Ti...	9.05	1545108	5	21	10	Finished Airing	Apr 29, 2019 to Jul 1, 2019	Spring 2019	['Production I.G', 'Dentsu', 'Mainichi Broadcast...] Funimation	Wit Studio	Manga	23 min. per ep.	R - 17+ (violence & profanity)	



05.

01.

DUPLICATE VALUES

```
data.duplicated().sum()
```

```
1631
```

```
data.drop_duplicates(keep='first', inplace=True)
data.shape
```

```
(8369, 15)
```

02.

NAN VALUES

```
nan_count_per_column = data.isna().sum()
print(nan_count_per_column)
```

Premiered	4359
Producers	0
Licensors	0
Studios	0
Source	0
Duration	0
Rating	17

17 rows in Rating should be eliminated, this represents 0.2% of our sample:

```
data_cleaned = data.dropna(subset=['Rating'])
```

03.

DUPLICATE TITLES & RESET INDEX

```
data_cleaned = data_cleaned.drop_duplicates(subset='Title', keep='first')
```

```
data_cleaned.reset_index(drop=True, inplace = True)
```

```
data_cleaned.shape
```

```
(4987, 15)
```

data
preprocessing



06.

```
start_date = []
end_date = []

for i in range(len(data_cleaned)):
    aired_value = data_cleaned.iloc[i, 7]

    if isinstance(aired_value, str) and 'to' in aired_value:
        s, e = aired_value.split('to')
    else:
        s = aired_value if isinstance(aired_value, str) else None
        e = aired_value if isinstance(aired_value, str) else None

    start_date.append(s.strip() if s else None)
    end_date.append(e.strip() if e else None)
```

convert object
to datestamp

Now we proceed to convert the data into a useful format (datestamp)

```
date_df = pd.DataFrame({'Release date': start_date, 'Completed date': end_date})

# Assigning the new columns
data_cleaned['Release date'] = date_df['Release date']
data_cleaned['Completed date'] = date_df['Completed date']

# Convert 'Release date' and 'Completed date' to datetime64 and set time to midnight
data_cleaned['Release date'] = pd.to_datetime(data_cleaned['Release date'], errors='coerce').dt.normalize()
data_cleaned['Completed date'] = pd.to_datetime(data_cleaned['Completed date'], errors='coerce').dt.normalize()
```

final dataset

```
data_cleaned_2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4964 entries, 0 to 4963
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Title             4964 non-null    object  
 1   Score              4964 non-null    float64 
 2   Vote               4964 non-null    int64   
 3   Ranked              4964 non-null    int64   
 4   Popularity         4964 non-null    int64   
 5   Episodes            4941 non-null    Int64  
 6   Status              4964 non-null    object  
 7   Aired               4964 non-null    object  
 8   Premiered           2431 non-null    object  
 9   Producers            4964 non-null    object  
 10  Licensors           4964 non-null    object  
 11  Studios              4964 non-null    object  
 12  Source               4964 non-null    object  
 13  Duration             4964 non-null    object  
 14  Rating               4964 non-null    object  
 15  Release date         4964 non-null    datetime64[ns]
 16  Completed date       4922 non-null    datetime64[ns]
 17  Release Year          4964 non-null    int32  
dtypes: Int64(1), datetime64[ns](2), float64(1), int32(1), int64(3), object(10)
memory usage: 683.6+ KB
```

Highest/Lowest Score



```
highest_score_anime = data_cleaned_2[data_cleaned_2['Score'] == data_cleaned_2['Score'].max()]
lowest_score_anime = data_cleaned_2[data_cleaned_2['Score'] == data_cleaned_2['Score'].min()]
```

```
print("Highest Scored Anime:")
print(highest_score_anime[['Title', 'Score']])
```

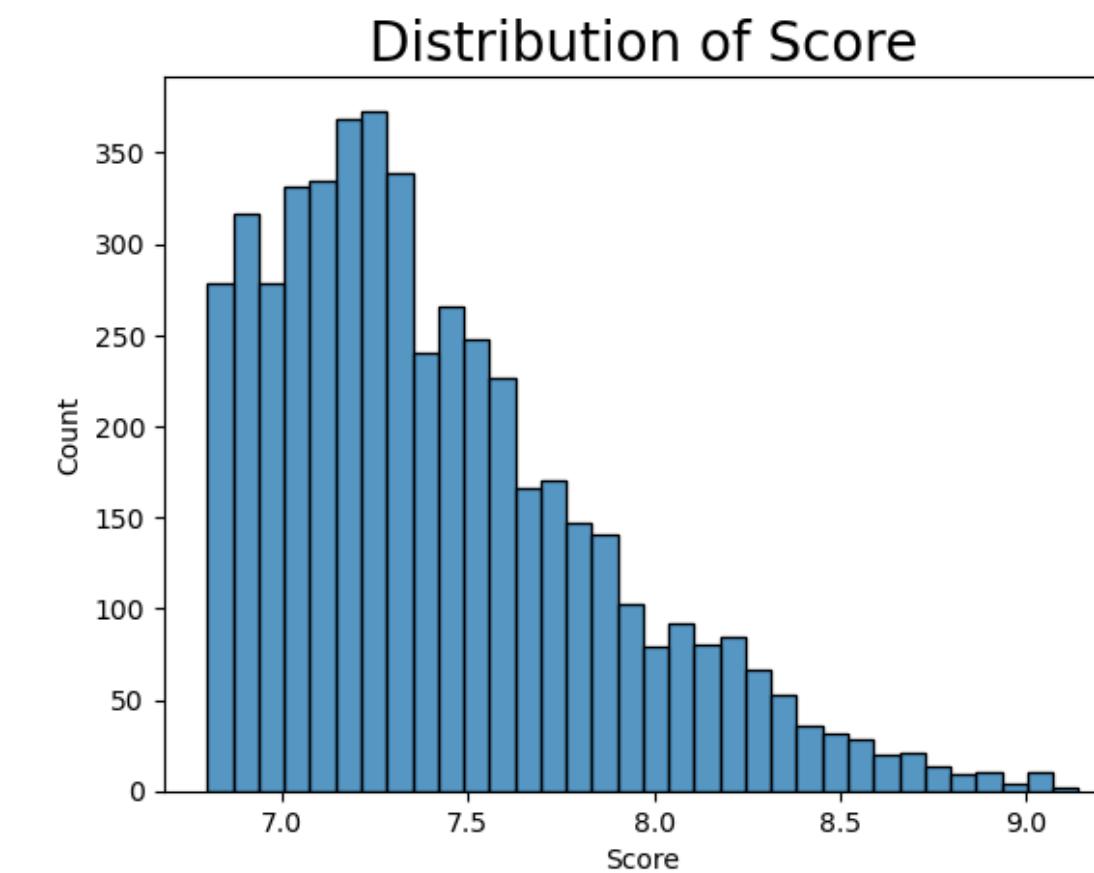
```
print("\nLowest Scored Anime:")
print(lowest_score_anime[['Title', 'Score']])
```

Highest Scored Anime:

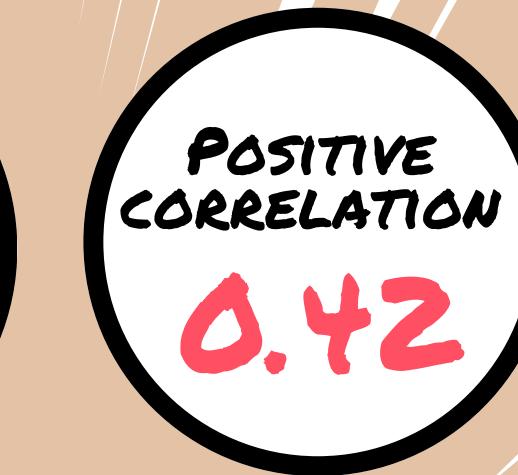
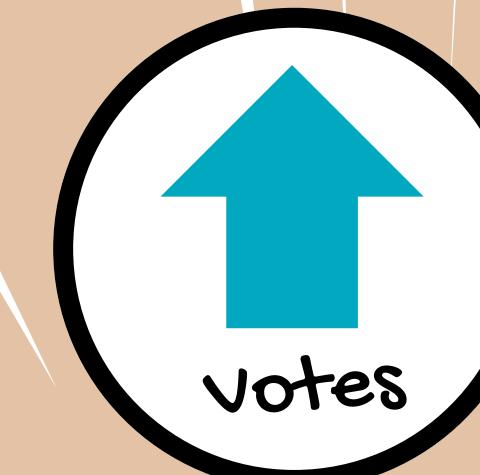
	Title	Score
0	Sousou no FrierenFrieren: Beyond Journey's End	9.14

Lowest Scored Anime:

	Title	Score
4922	Lupin III: Bye Bye Liberty - Kiki Ippatsu! Lupi...	6.8



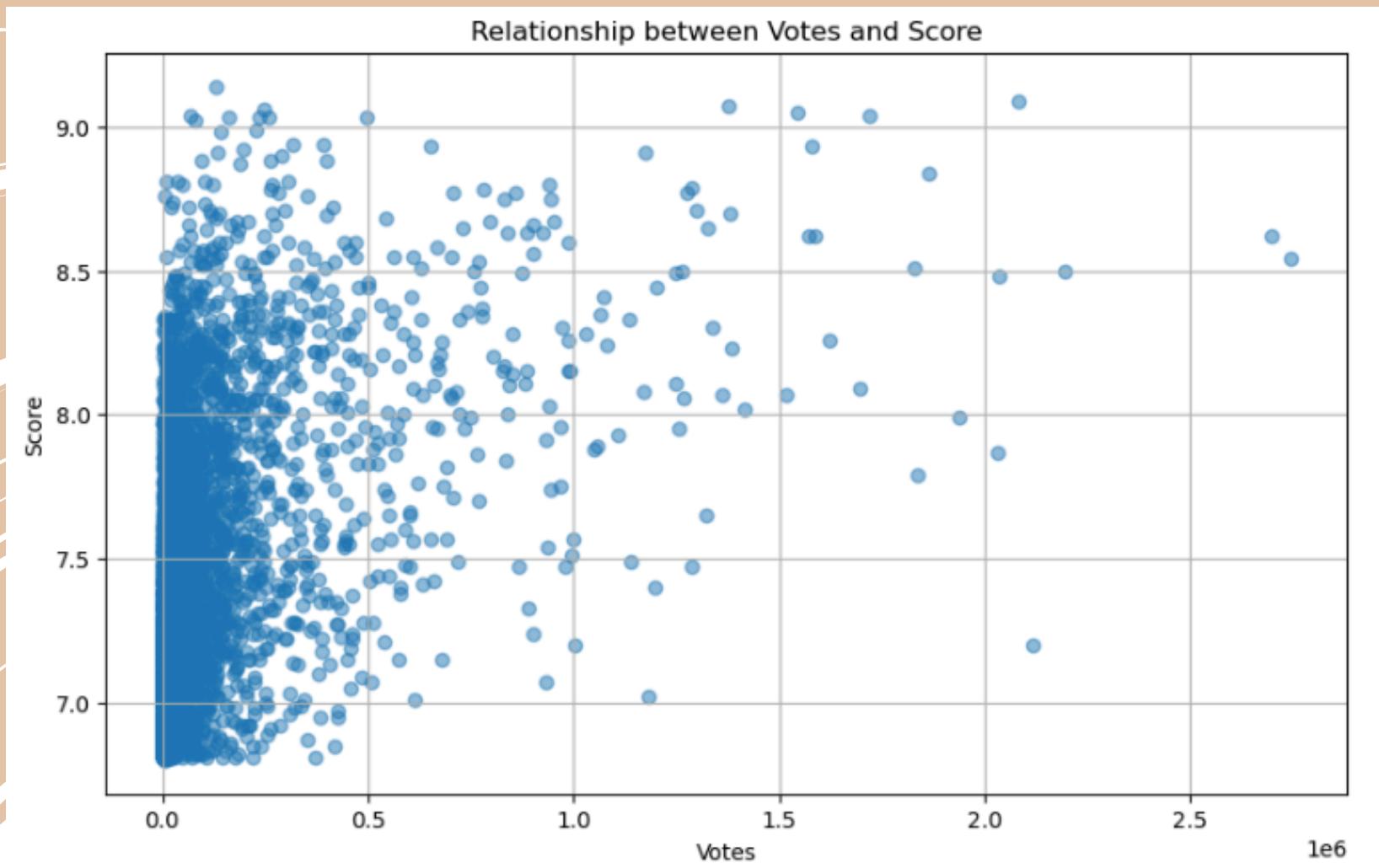
Score vs Votes



This suggests a potential relationship between anime with higher engagement are perceived as higher quality

POSSIBLE REASONS

- Popular anime are well received by the audience.
- High score animes attract more viewers (and more votes).

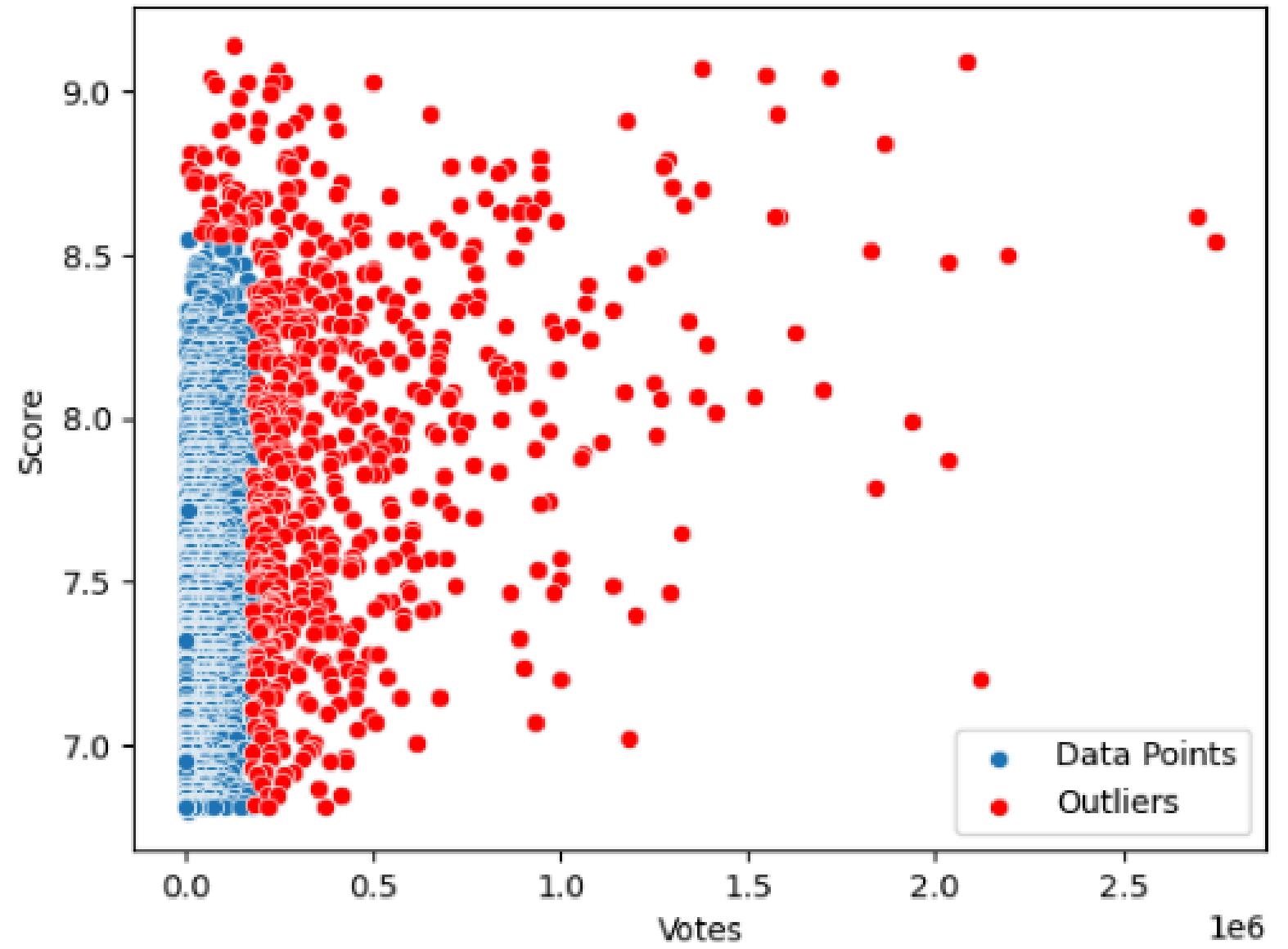


10.

outliers

Boxplot of Anime Scores

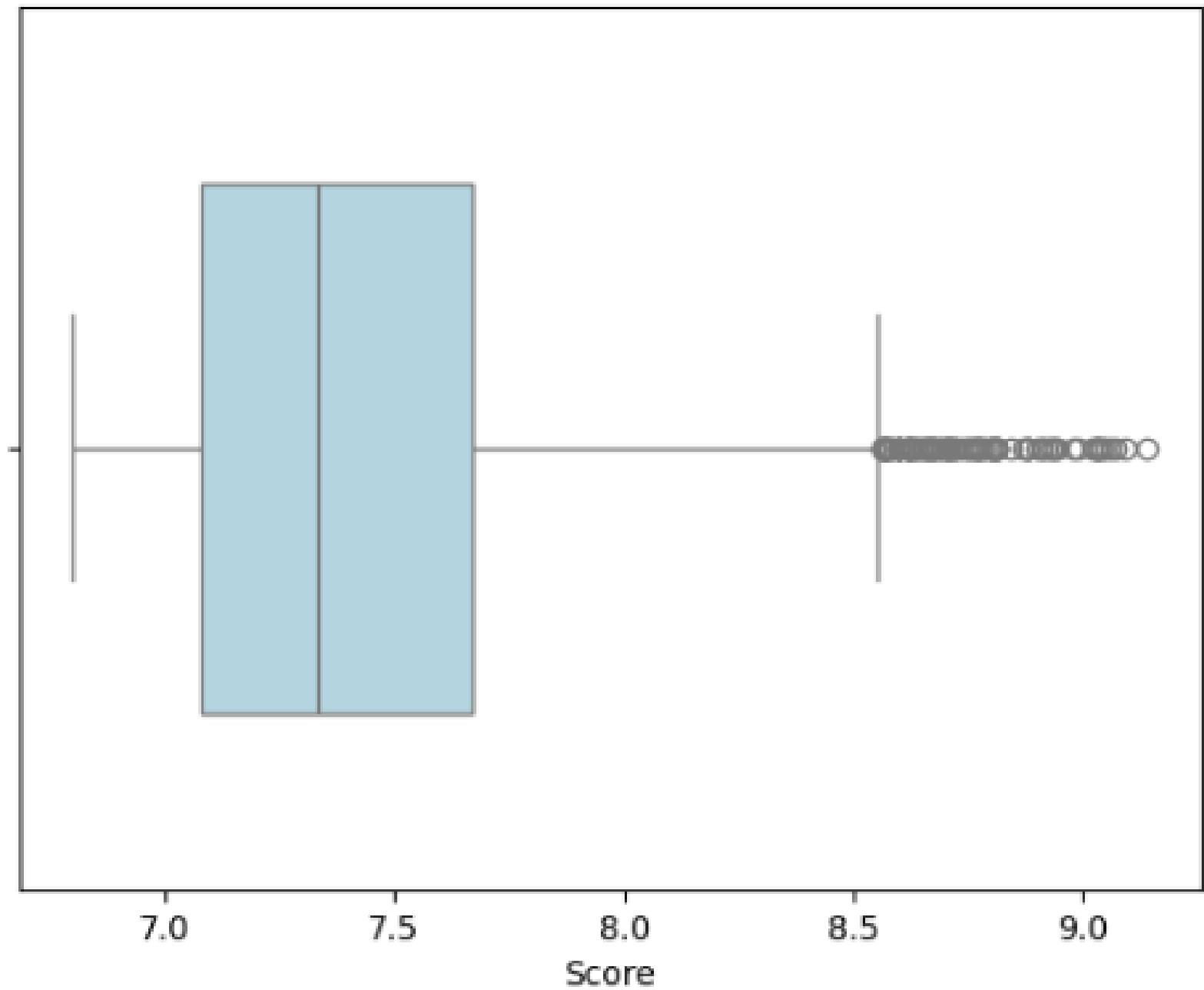
Scatter Plot with Outliers



- **High Votes, Low Score:** Anime with a large number of votes but low score.

Possible reasons:

- Heavily marketed anime that failed to meet expectations.
- Sensitive content that led to mixed reviews.



- **Low Votes, High Score:** An anime with very few votes but an unusually high score.

Possible reasons:

- An underrated anime appreciated by a small, dedicated fanbase.
- Recent release or limited accessibility, resulting in fewer votes.



high number of
votes but low
Score

There are animes with many votes and low scores, but there is no case that an Anime has a low score with a great amount of votes.

```
votes_limits = only_series_3['Vote'].quantile(0.75)
score_limits = only_series_3['Score'].quantile(0.25)

high_votes_low_score = only_series_3[(only_series_3['Vote'] > votes_limits) & (only_series_3['Score'] < score_limits)]

print("Anime with high Votes but low Score:")
print(high_votes_low_score[['Title', 'Vote', 'Score']])
```

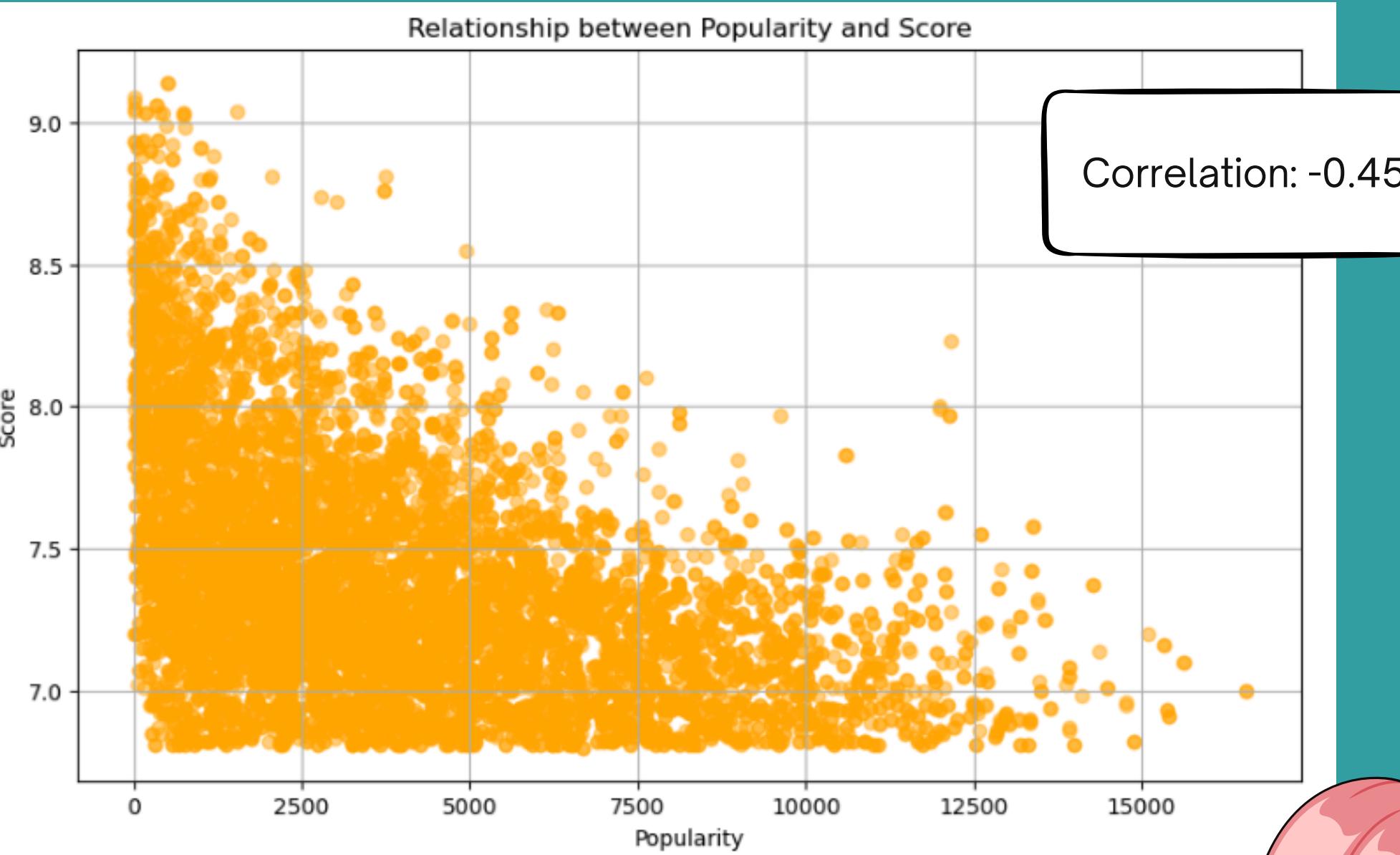
Anime with high Votes but low Score:

		Title	Vote	Score
3587	Madan no Ou to Vanadis	Lord Marksman and Vanadis	177943	7.11
3630		Love Hina	135297	7.10
3669		Himouto! Umaru-chan	378005	7.10
3680		Black Bullet	484966	7.09
3698		Vampire Knight Guilty	222405	7.09
...	
4879	Gakusen Toshi Asterisk	The Asterisk War	372608	6.81
4906		Kiss x Sis	182980	6.82
4934	Shinmai Maou no Testament Burst	The Testament o...	217845	6.81
4939	Black★Rock Shooter (TV)	Black Rock Shooter	175137	6.81
4960	Taimadou Gakuen 35 Shiken Shoutai	Anti-Magic Ac...	146677	6.81



Popularity vs Score

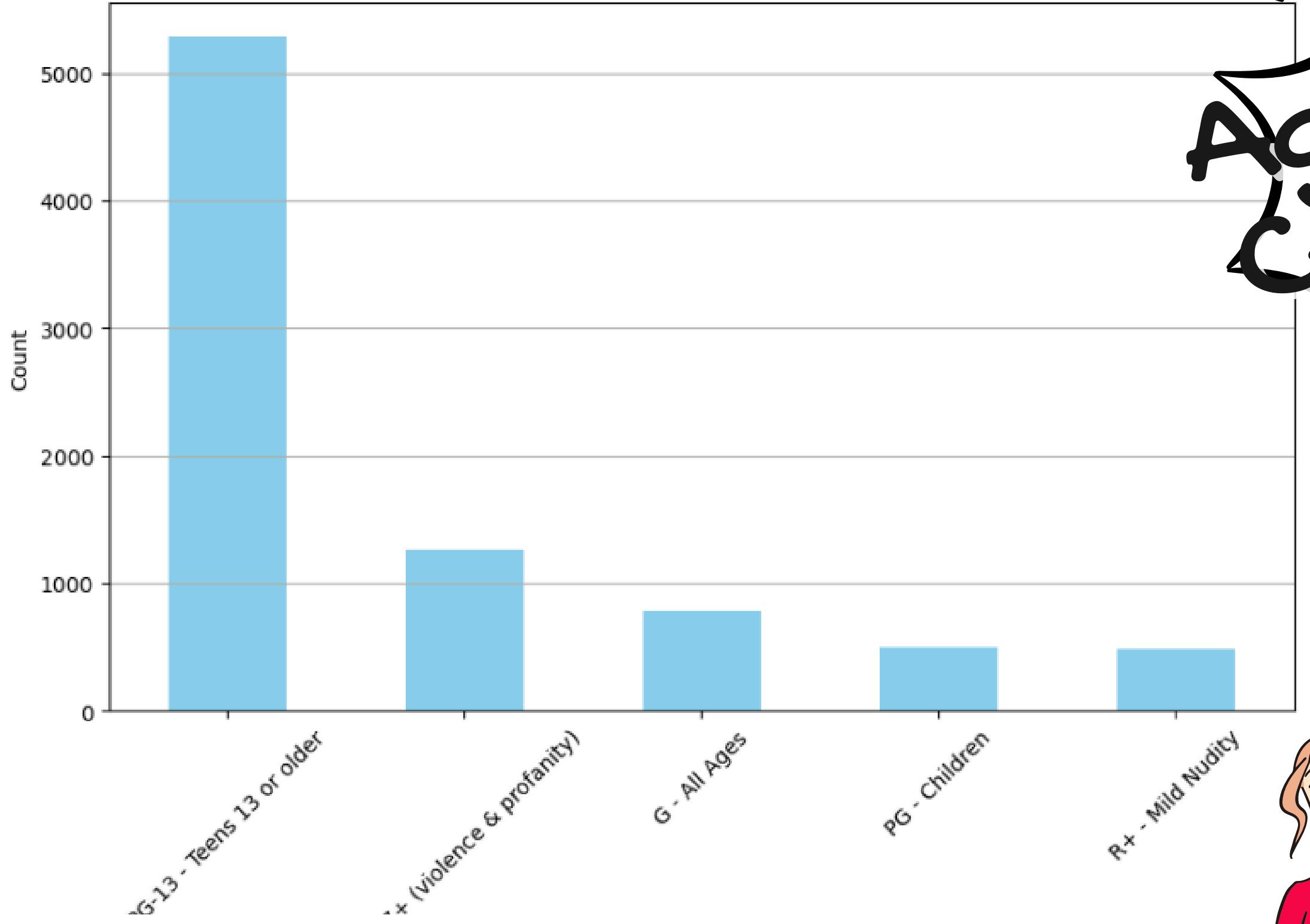
A highly rated anime (higher Score) often becomes more popular, resulting in a low Popularity rank. Conversely, anime with lower Scores are less likely to achieve top popularity rankings.



Popular anime tends to be high quality: This could indicate that audiences generally favor and vote for anime that are critically acclaimed or well-made.



Most Common Age Rating Categories



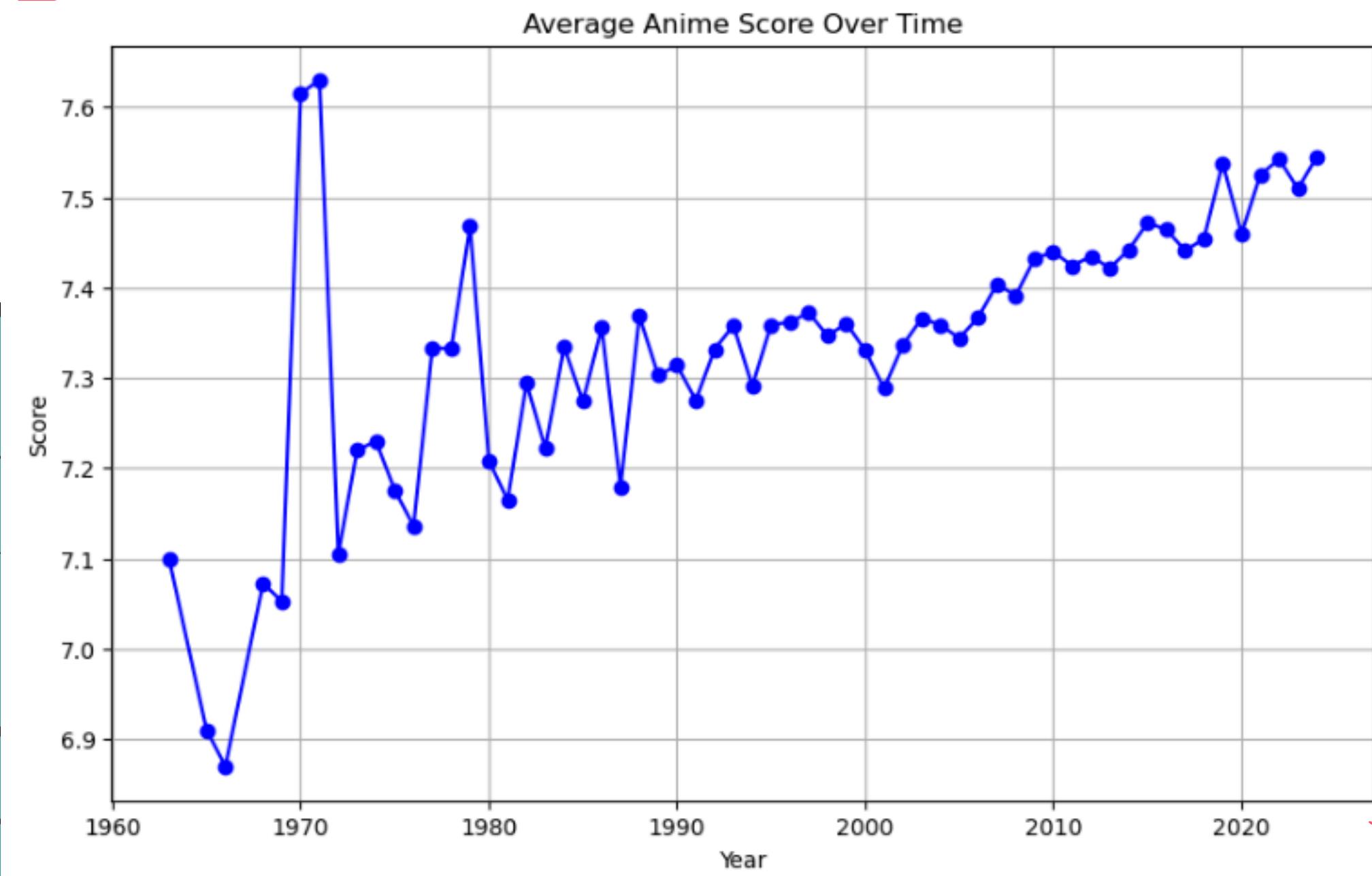
Age Rating Categories



SCORE EVOLUTION

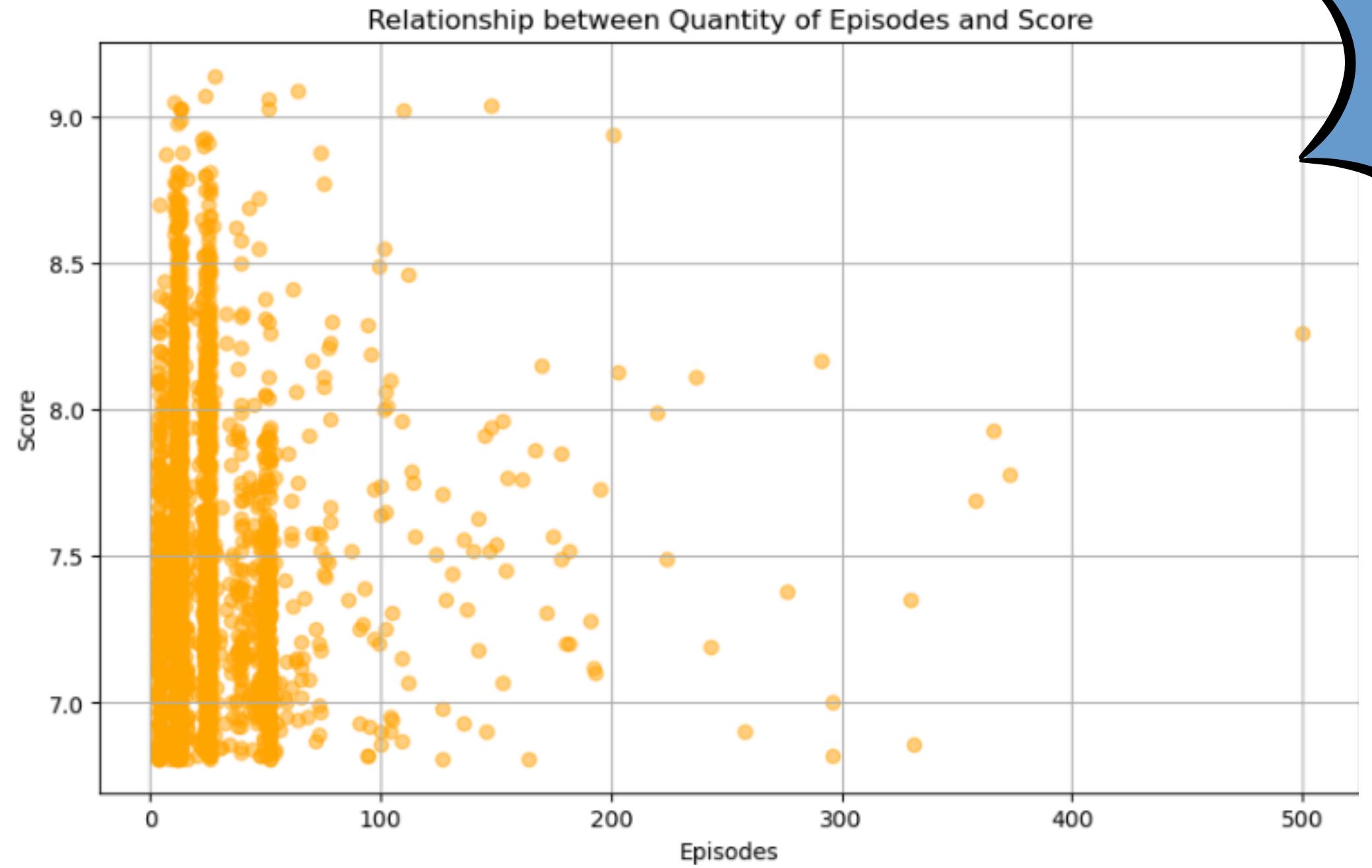
OVERTIME

This graph suggests that there is a slight upward trend in the average score over the years.



15.

Episodes vs Score

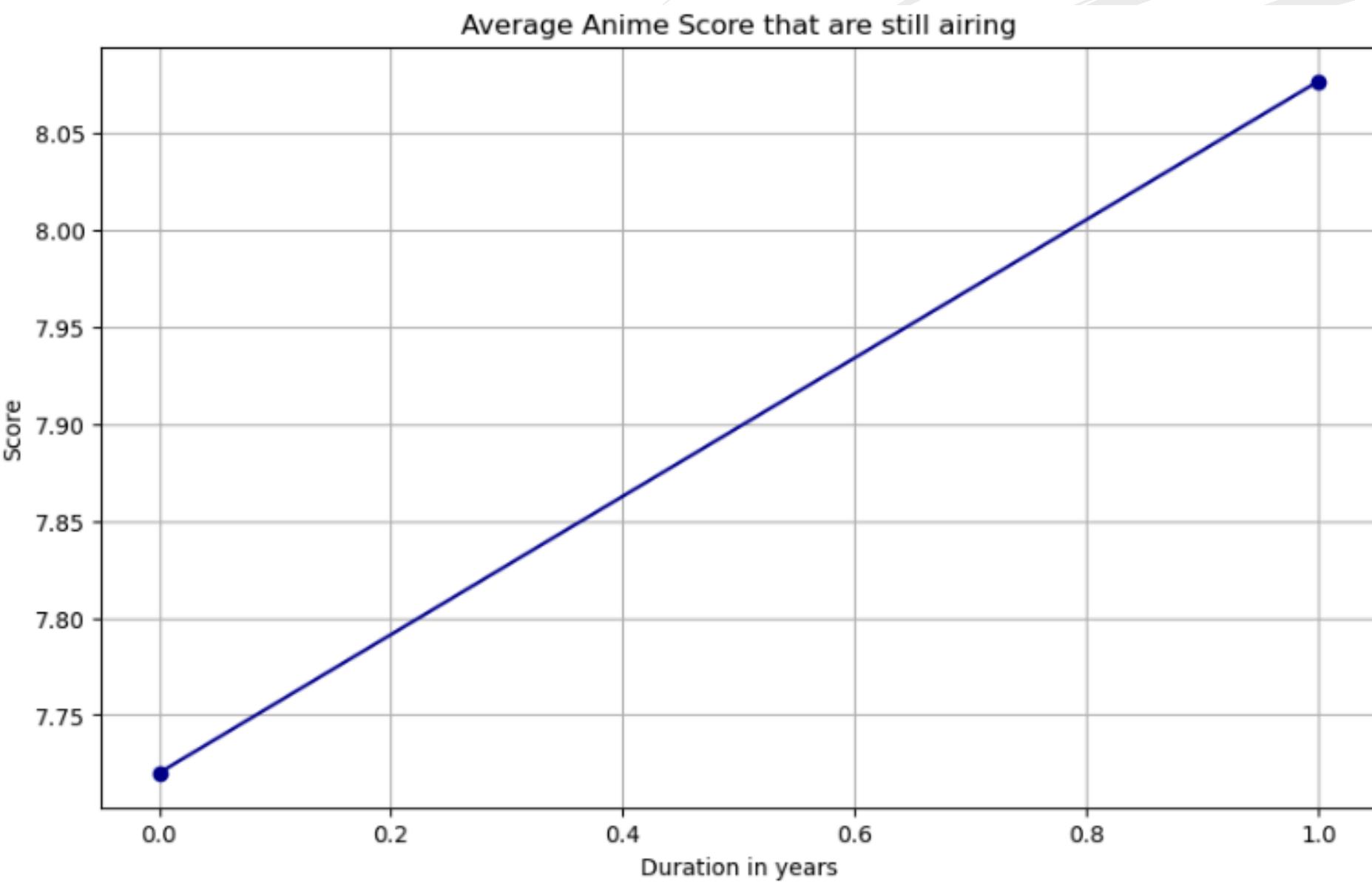
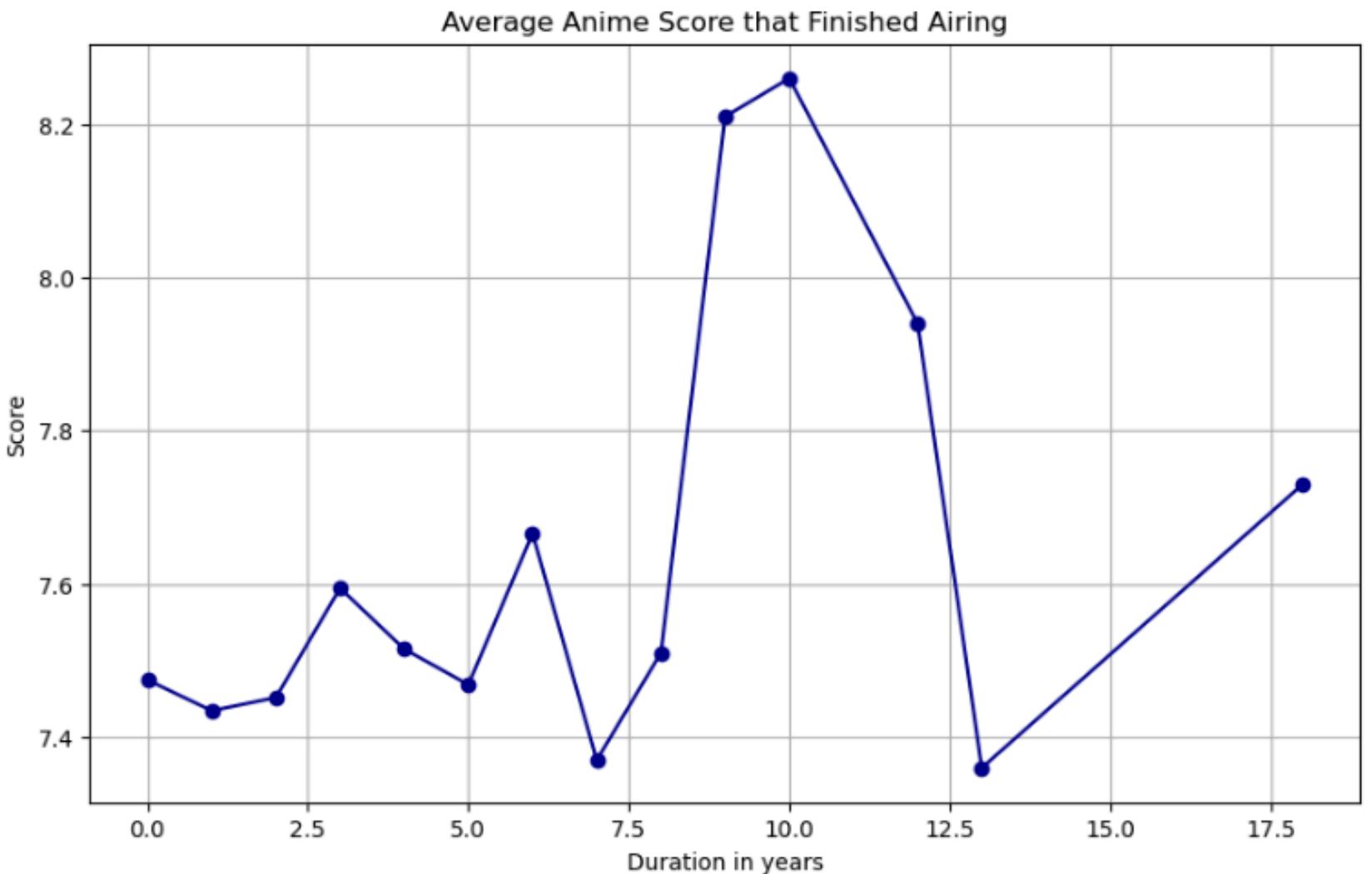


The correlation is negative and near zero (not very strong). However this means that the variables are inversely proportional, meaning that the higher the score, the less the quantity of Episodes.



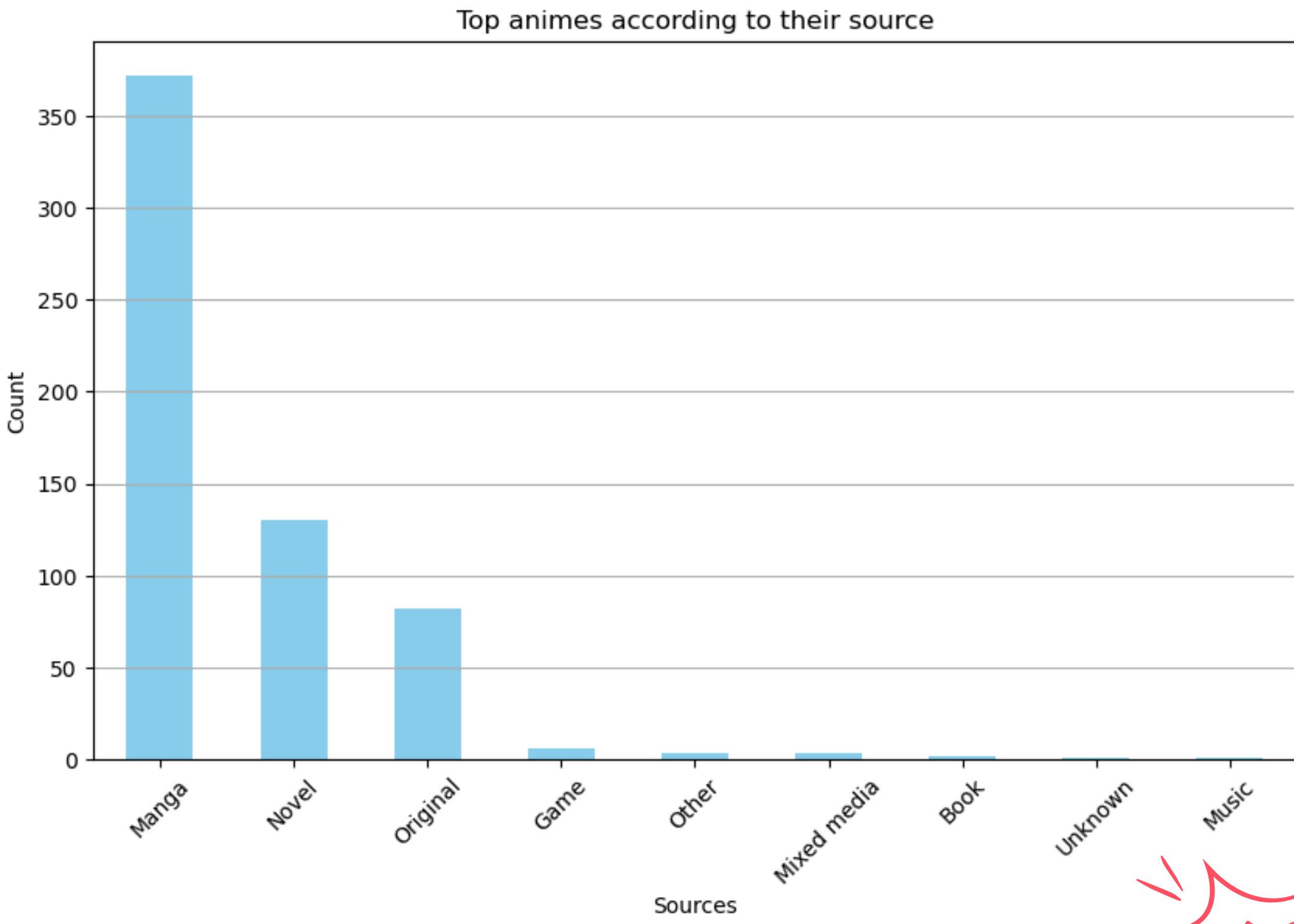
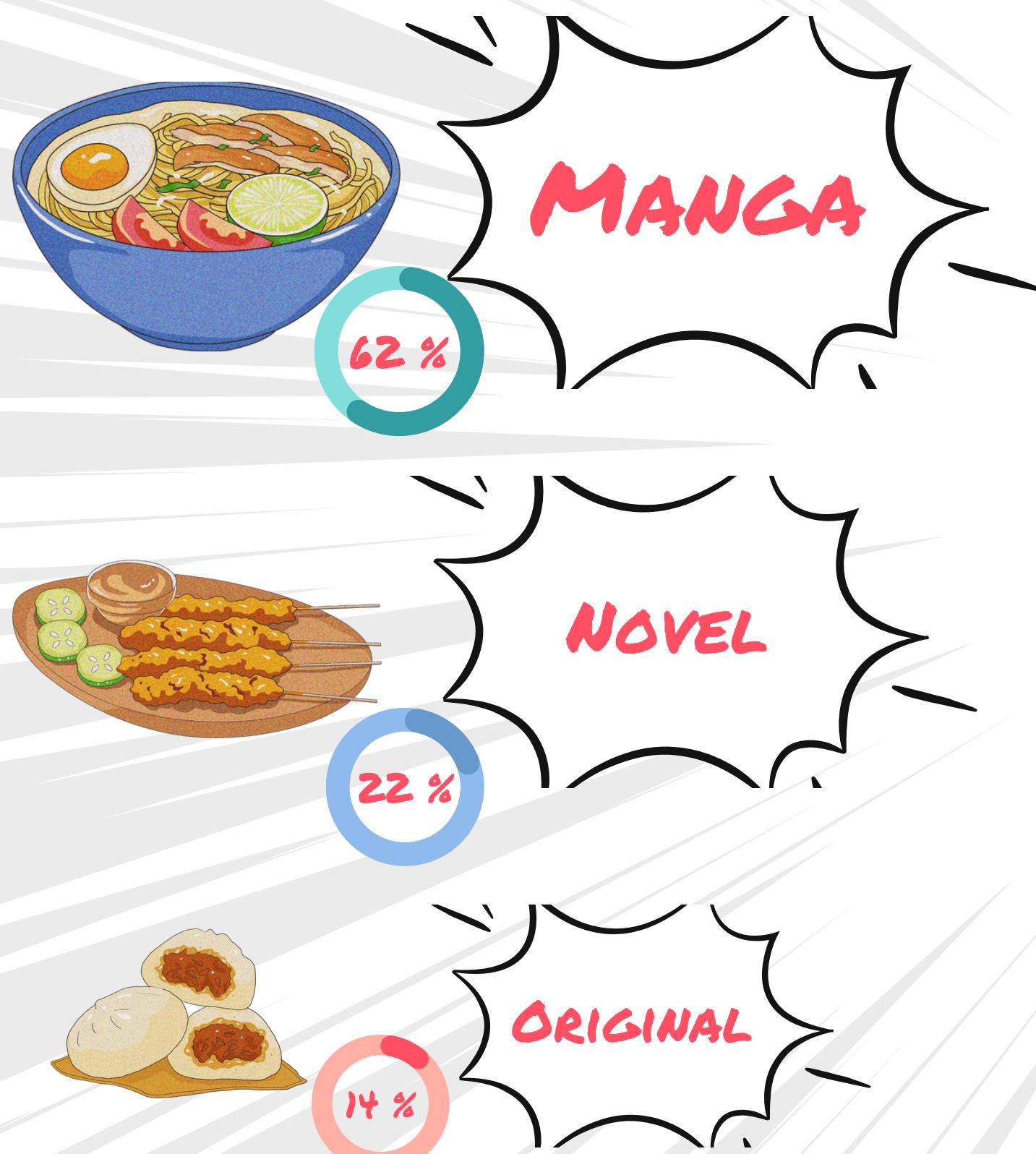
Status vs Score

In general terms Animes that already finished airing have an average score between 7.4 and 7.7. However, there are some exceptional cases, where animes that were on air between 7.5 and 12.5 years hit higher peaks (7.9 - 8.3)



Animes that are still airing have an increasing trend in the score results, starting above the average (7.7 - 8.1). It is expected to have a growing trend in this case, because as the series is still developing the score is not definitive and also, as there are less animes in comparison with the ones that finished airing, the score tends to be higher.

most common Sources



Average
duration per
episode

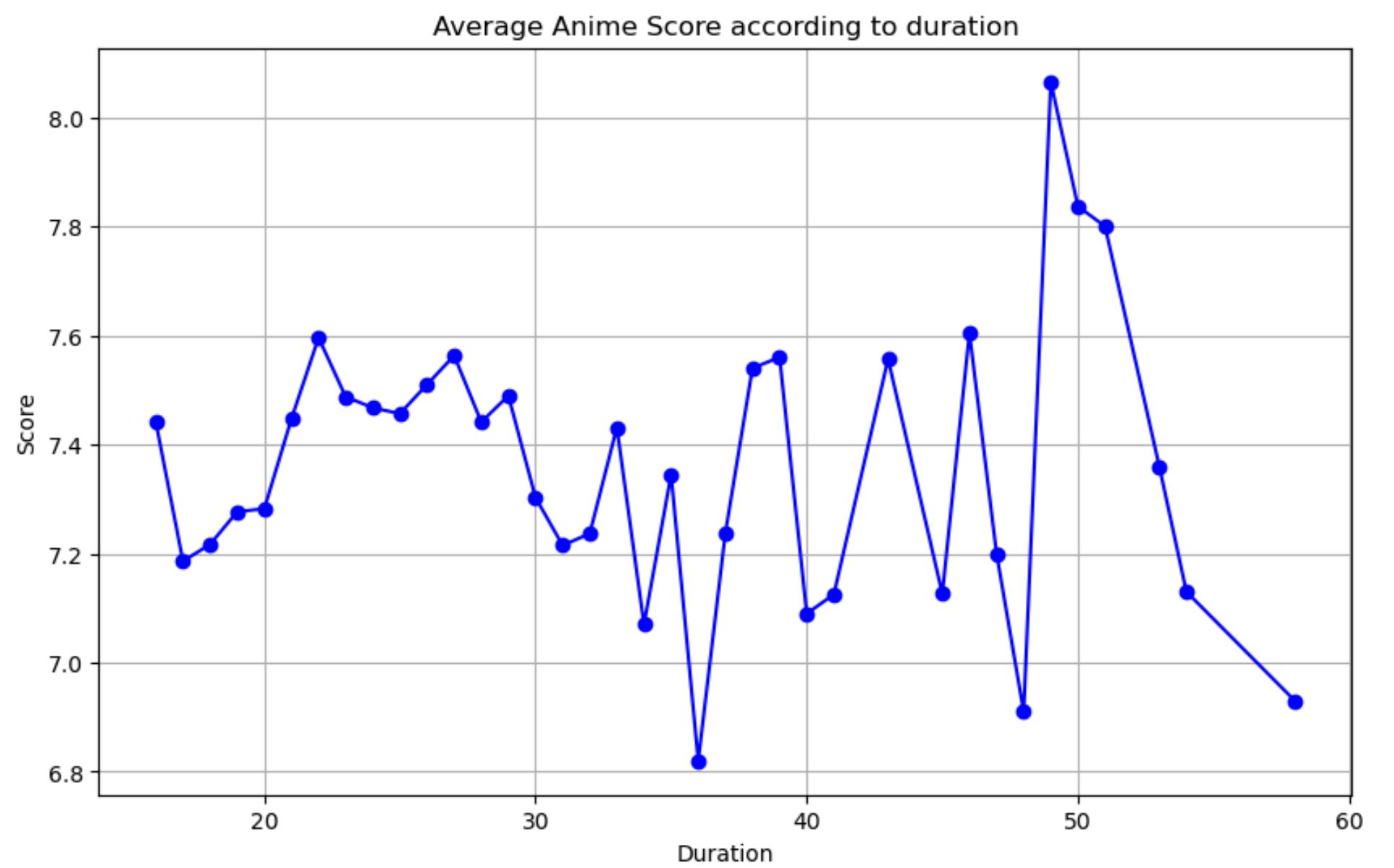
24 MINUTES



The range is
between 22 - 26
min per episode



Duration vs Score



In general terms, animes last between 20 and 30 minutes have average scores. However, there are exceptional cases where Episodes that lasted longer had a better score. From this, we can say that there is no specific trend regarding the duration of the episode and the Score.

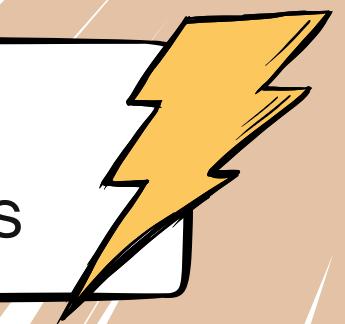
20



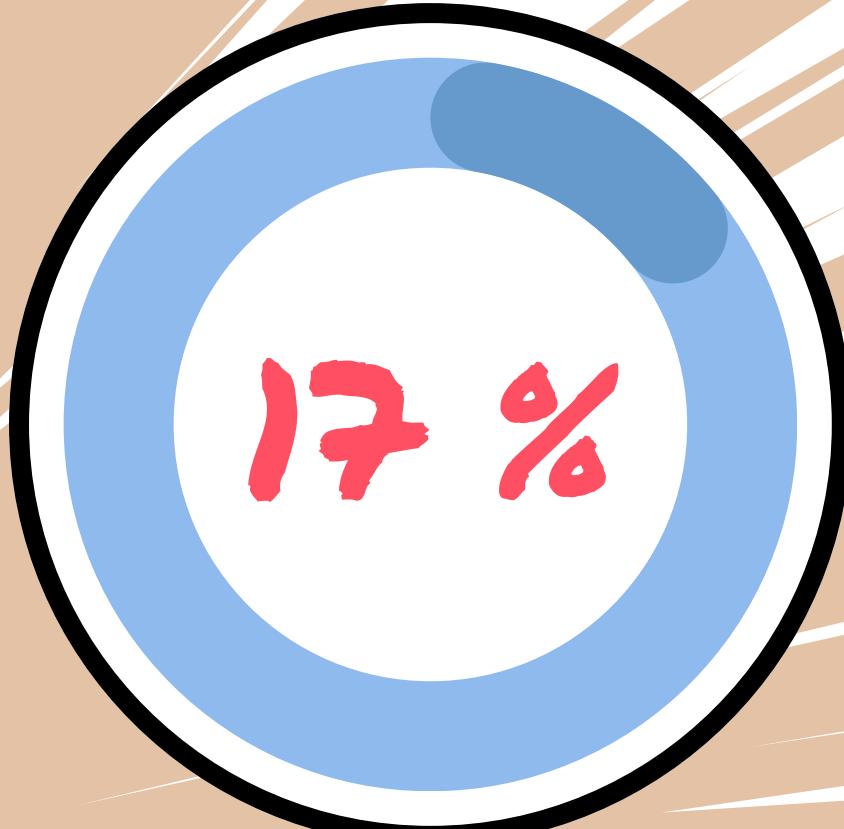
Amount
of
Episodes



12-13
Episodes



24-26
Episodes



Mixed
Results

21

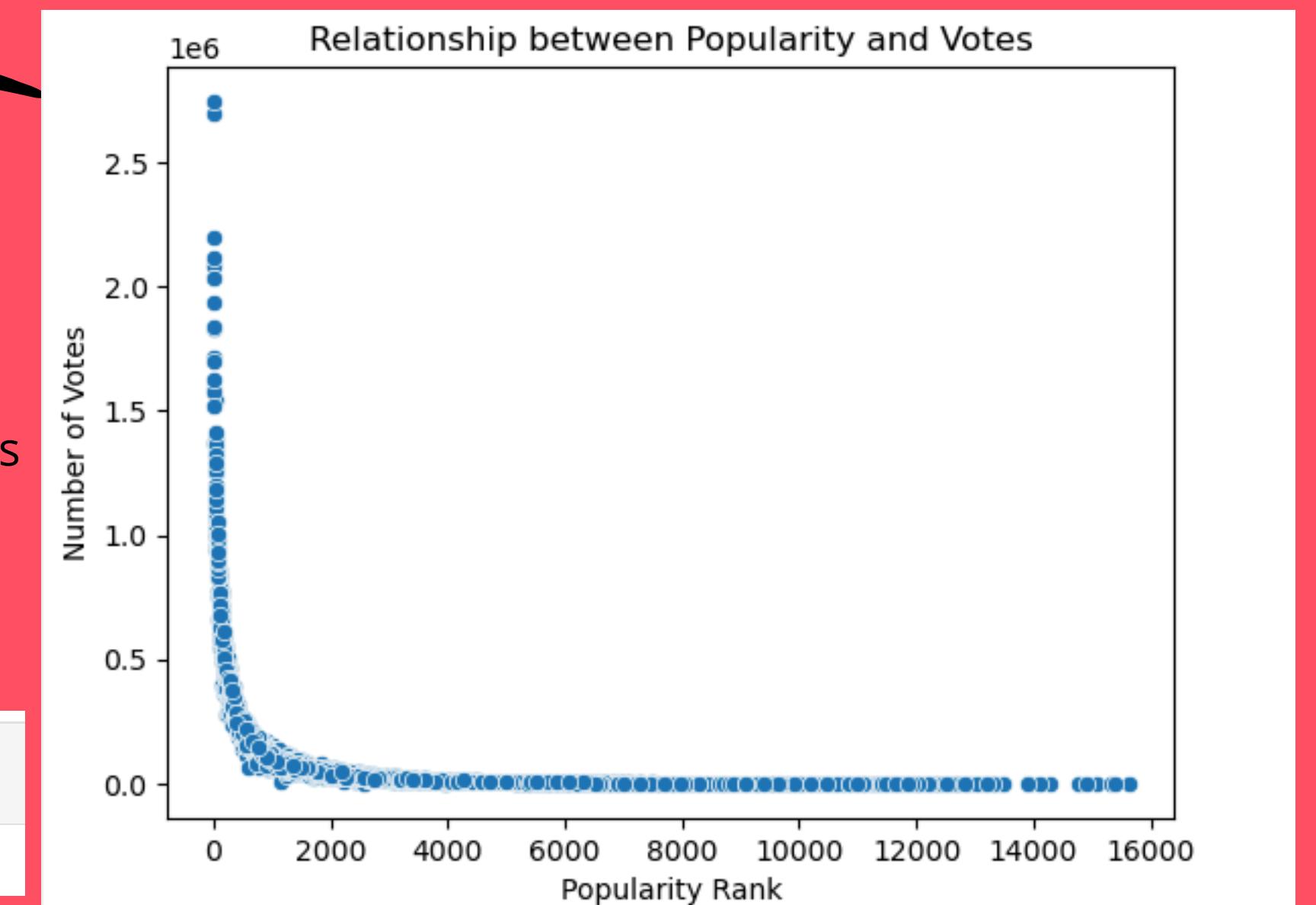
Popularity vs Votes

This relationship shows that more popular anime (lower rank) tend to gather a higher number of Votes.

A negative correlation would indicate this trend.

```
corr_pop_vote = only_series_3['Popularity'].corr(df['Vote'])  
print("Correlation between Popularity and Votes:", correlation)
```

```
Correlation between Popularity and Votes: -0.44121418862990663
```



Conclusions

01.

The higher the votes,
higher the score

02.

The higher the score, less
quantity of Episodes

Current airing animes have
better score than animes
that finished airing

03.

Most common Rating
PG -13

04.

Most common
Sources: Manga, Novel
and Original stories

Most episodes last 24
minutes and have an
average score around
7.5.



THANK YOU!



Drop your
questions here