Data Analytics Project

Customized Anime Recommendations for Users

Presented by: Maithreyee B Bharadwaj
Email: maithreyee95@gmail.com
Date: 25 July 2023

Problem Statement

I am a junior data analyst working for a company called Myanime, a website which contains a variety of animes. My manager asked me to give recommendations to users based on what they have watched and new users the top animes based on a variety of criteria. My job is to figure out a data driven strategy to give recommendations.

Q. What is an anime?

A. Anime (アニメ) is a Japanese term for hand-drawn or computer animation. The word is the abbreviated pronunciation of "animation" in Japanese, where this term references all animation. Outside Japan, anime is used to refer specifically to animation from Japan or as a Japanese-disseminated animation style.

Exploring the database

Exploration of the database was split into three parts

1. Cleaning the database
2. Analysing the database
3. Visualizations

Cleaning the database

1. The database has three excel files which are called anime, reviews and profile.
2. Each file has almost 30,000-40,000 records.
3. I had to check if I could automatically import the data into mysql workbench using the import wizard.
4. Here is where I ran into a lot of issues and with which I understood why cleaning of the database is so important.
5. The major issues I faced were :

    5.1. Special characters: If the excel file has a lot of special characters it does not get imported. Even importing manually by creating a table and importing the csv, the records get dropped. Here the excel had to be cleaned by replacing special characters with a blank space.

    5.2. I had to be very careful of the data types which was used to define the variables while creating tables.

    5.3 #Name? is an error when the names have an = or @ so such values had to be changed by replacing them with an underscore(_) as these records were getting dropped while importing.

Analysis

The database consists of three tables

Table	Rows	Columns
Anime	16,217	9
Review	192,117	10
Profile	81,727	3

Anime: This table consists of the anime name, id, genre, no of episodes, the dates from which it aired on television, the no of people who have watched the anime, popularity, score and the rank of each anime

Review: This table consists of the anime id, user id of the users who have scored a particular anime, score, and score on various categories which are story, sound, animation, characters, enjoyment

Profile: This table consists of the user id, the gender of each user and their favourite animes.

Visualization

Since our website provides a variety of animes, we have to customize the animes based on user preferences. I have grouped the anime based on some of the choices we are going to give our users.

The categories are as follows:

User based Categories	Anime based Categories
1. Popularity	***1.Story***
2. Score	***2.Animation***
3. Gender	***3. Enjoyment***
4. Genre	4.Characters
	5. Sound

MyAnime user page

MyAnime

baekbeans

Akatsuki no yona

Fullmetal Alchemist

Clannad -After stroy

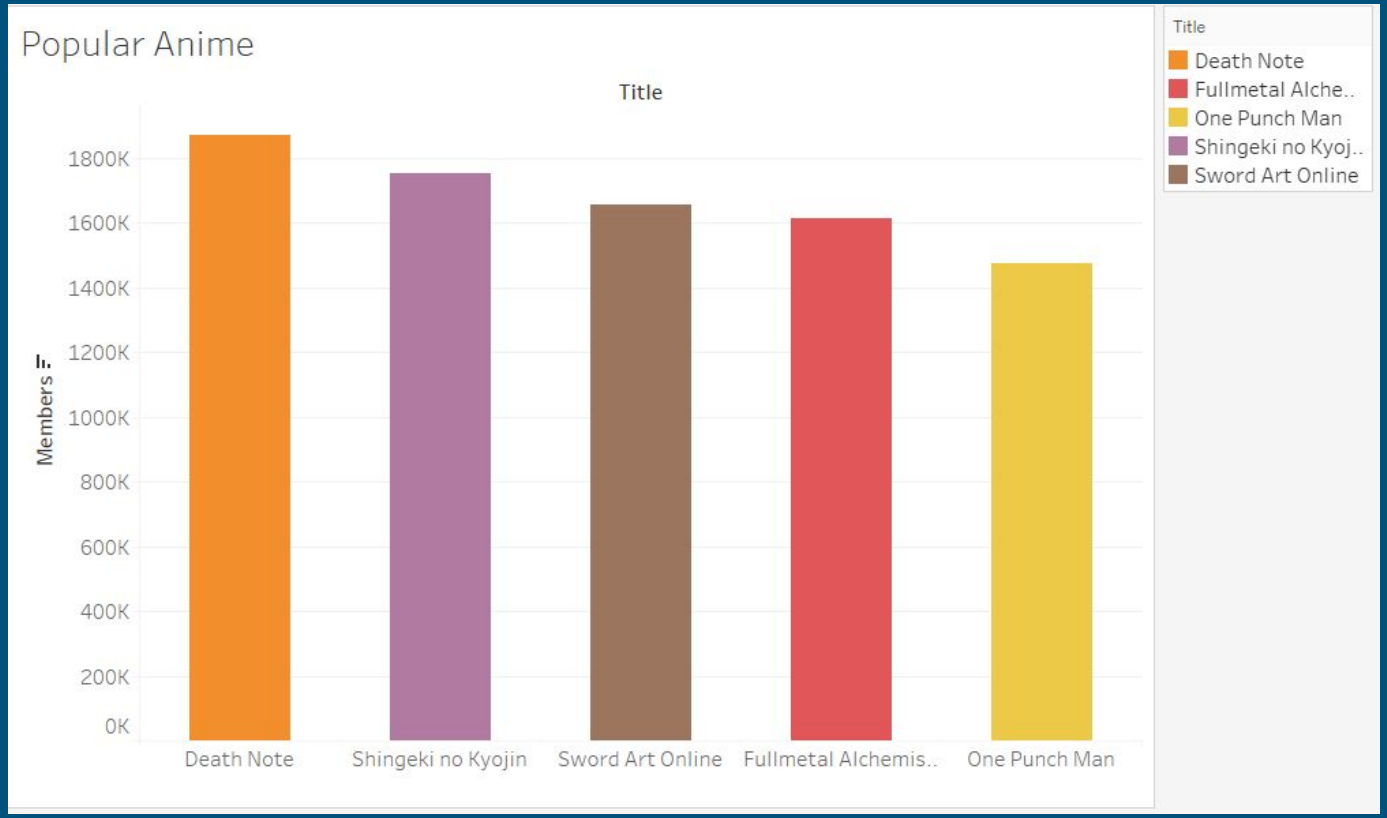
Steins Gate

Ouran high school host club

- This is the web page when a user logins into his/her account.
- As an example, I have taken a female user.
- These are the anime recommendations based on the results obtained through analysis of the database.
- Let me show you my analysis next.

I. User based categories

1.Popularity



Here the popularity of the anime is based on the number of users who have watched these animes

Anime	Users watched
Death Note	1,871,043
Shingeki no Kyojin	1,754,979
Sword Art Online	1,657,823
Fullmetal alchemist :Brotherhood	1,615,084
One Punch Man	1,475,219

2.Score

- The graph on the left hand side shows the percentage of scores for various anime.
- It can be observed that 22% of the anime having a score of 10.
- The percentage of each score gradually decreases.
- Further analysis have to be done to get more users to rate an anime.
- In order to get the top 5 animes with highest average score, it has to be rated by at least 100 users.

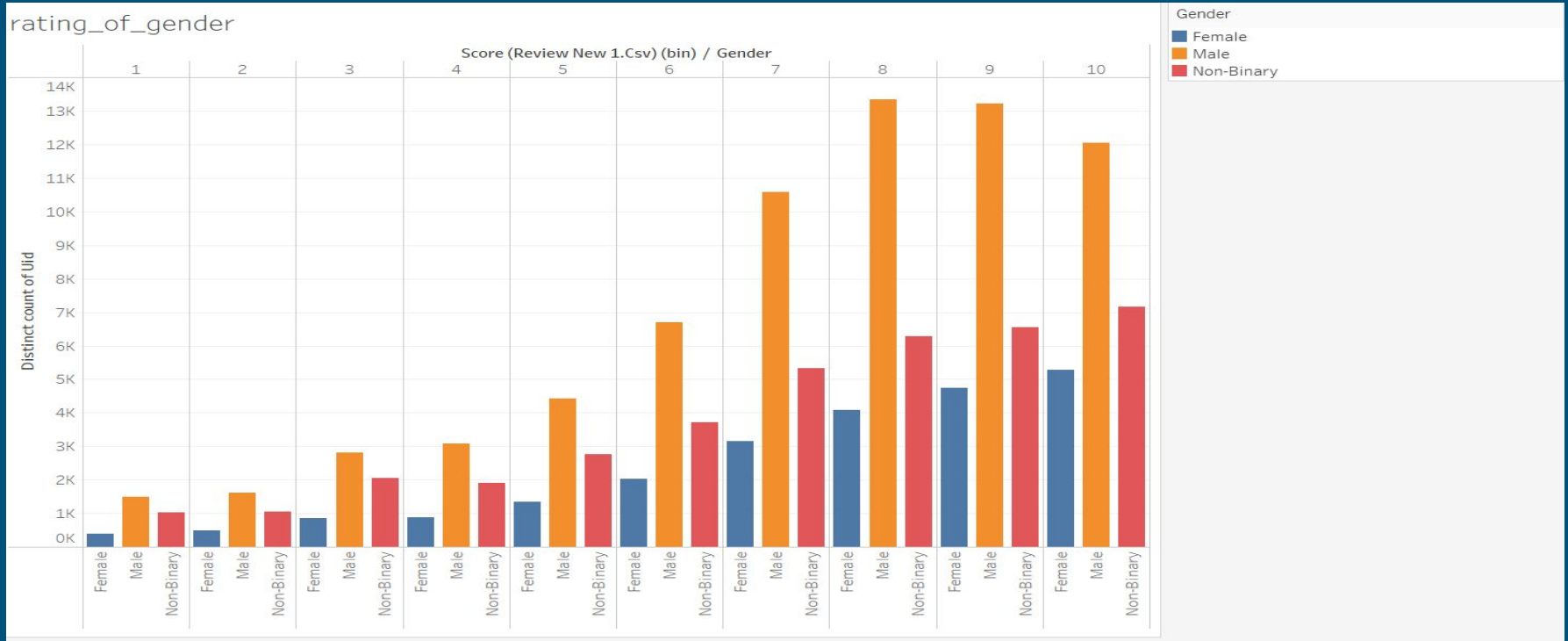
3. Gender

The three values under gender include:

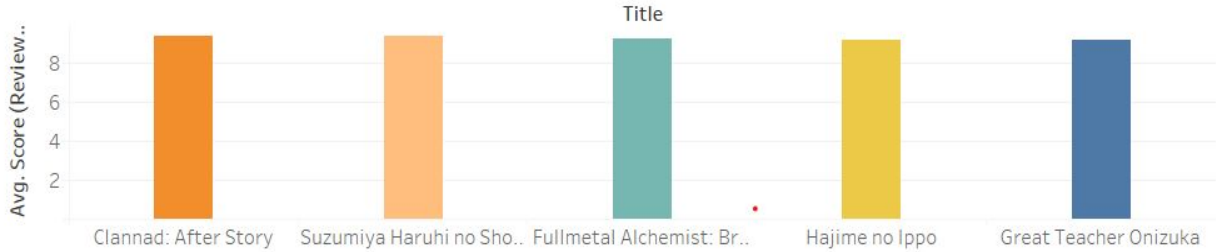
- Male
- Female
- Non-Binary

Below is a graph showing how each gender rates an anime.

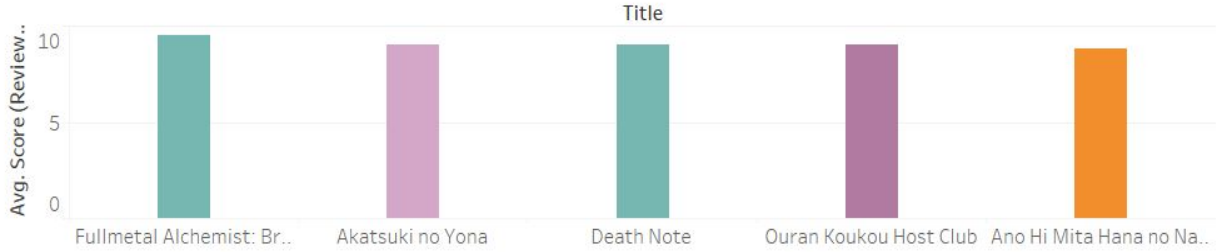
Insight:-It turns out that male users are more consistent in rating an anime than female users.



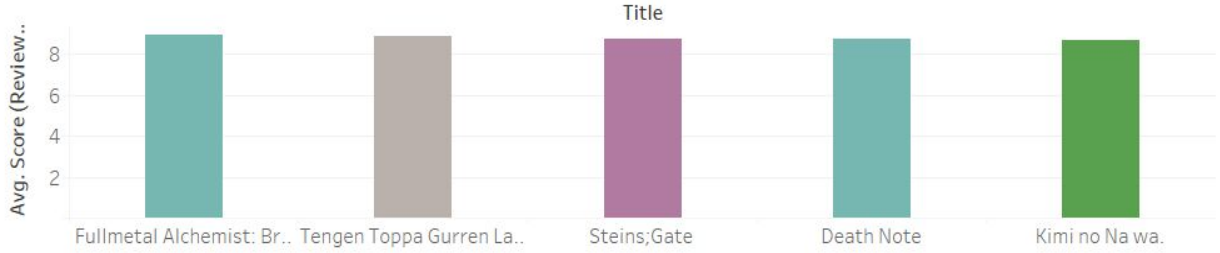
Male watched anime



Female_watched_anime



Gender_not_specified



These graphs show the top 5 most watched animes among the various genders.

The X-axis = Anime Title
The Y-axis = Avg Score

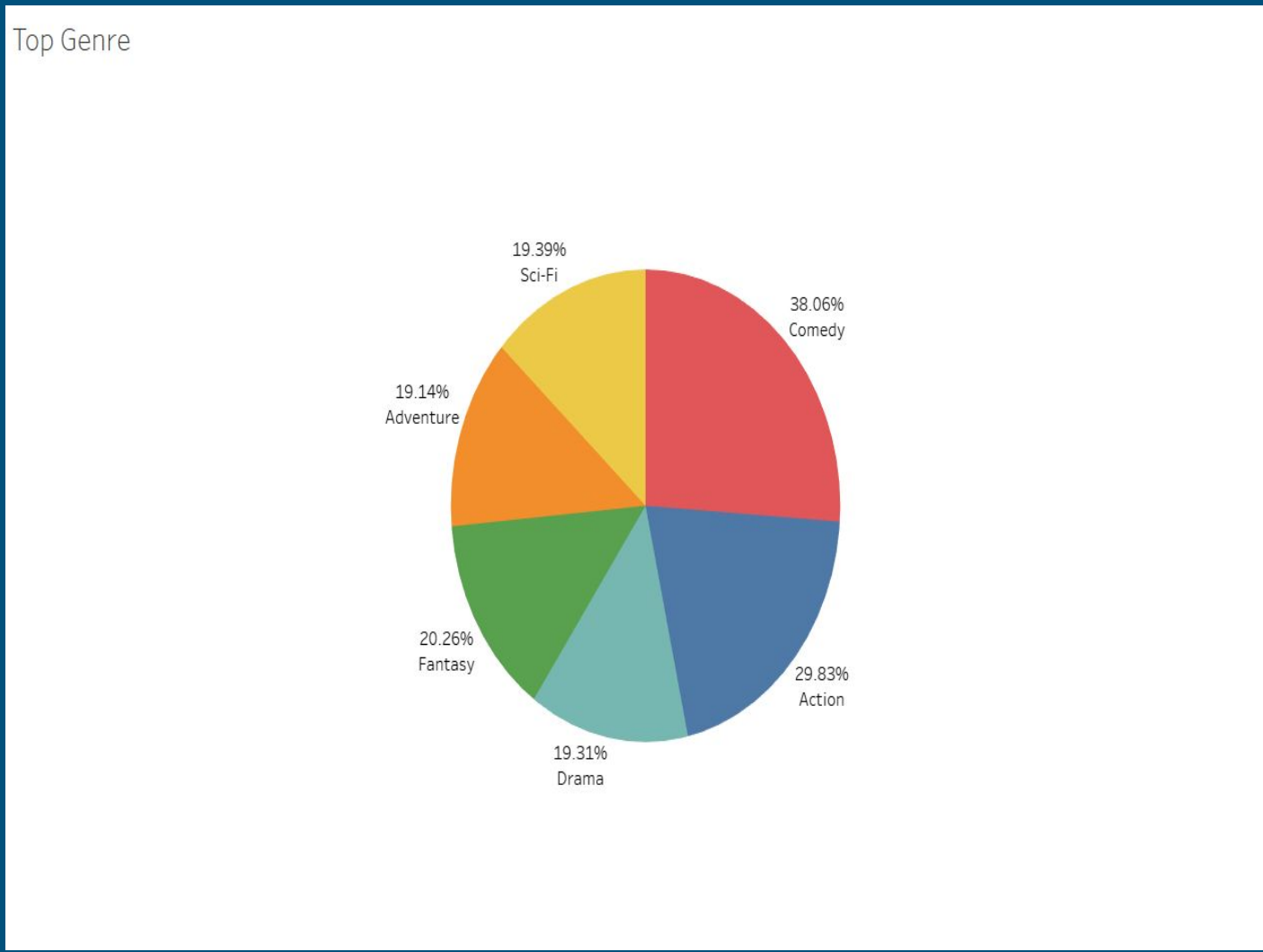
Condition -The number of users who have watched a particular anime and given a score should be more than 100.

Reason- To prevent animes which are watched by just 1 or 2 users and have been given the highest scores.

4.Genre

The top most watched genre are

1. Comedy
2. Action
3. Fantasy
4. Sci-Fi
5. Drama
6. Adventure



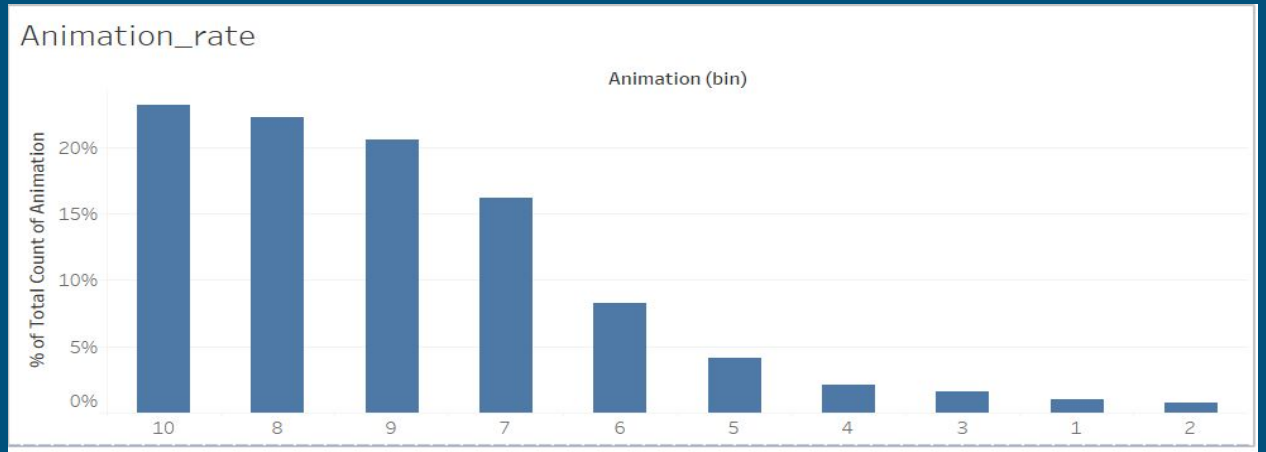
The top 3 animes from each genre

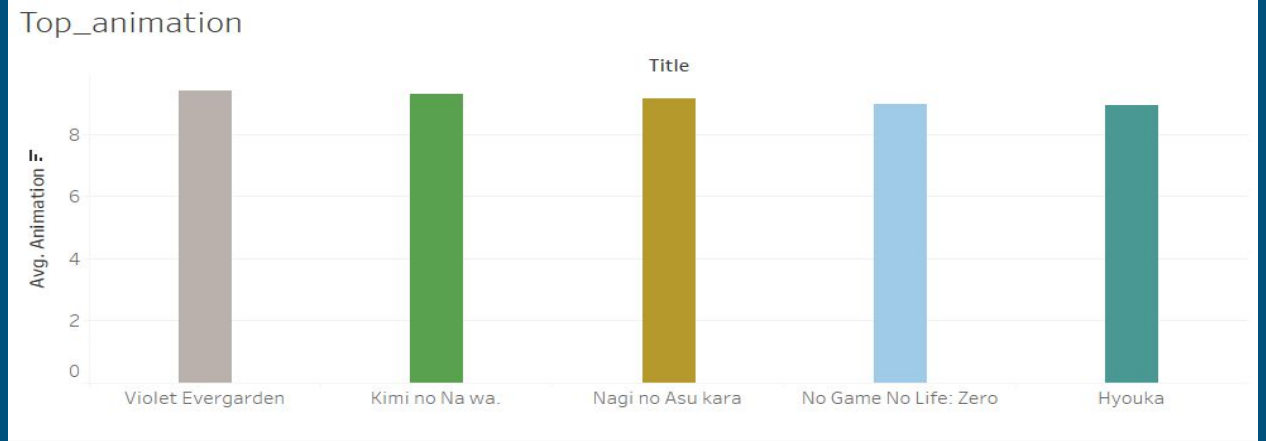
- Here is a table which contains the top three animes from each genre.
- An anime is picked for a particular genre if more than 100 users have watched it and have given a score.
- In this way an anime which has been watched by just one user and has been given the highest score does not get included.

Genre	Anime 1	Anime 2	Anime 3
Comedy	Suzumiya Haruhi no shoushitsu	Ansatsu Kyoushitsu 2nd Season	Fullmetal alchemist : Brotherhood
Action	Ansatsu Kyoushitsu 2nd Season	Fullmetal alchemist :Brotherhood	Gintama
Fantasy	Fullmetal alchemist :Brotherhood	Fullmetal alchemist	Ookami to Koushinryou
Sci-Fi	Suzumiya Haruhi no shoushitsu	Gintama	Steins Gate
Drama	Fullmetal alchemist :Brotherhood	Hajime no Ippo	Clannad After story
Adventure	Fullmetal alchemist :Brotherhood	Sen to chihiro no kamikakushi	Sora yori mo tooi Basho

II.Anime based categories

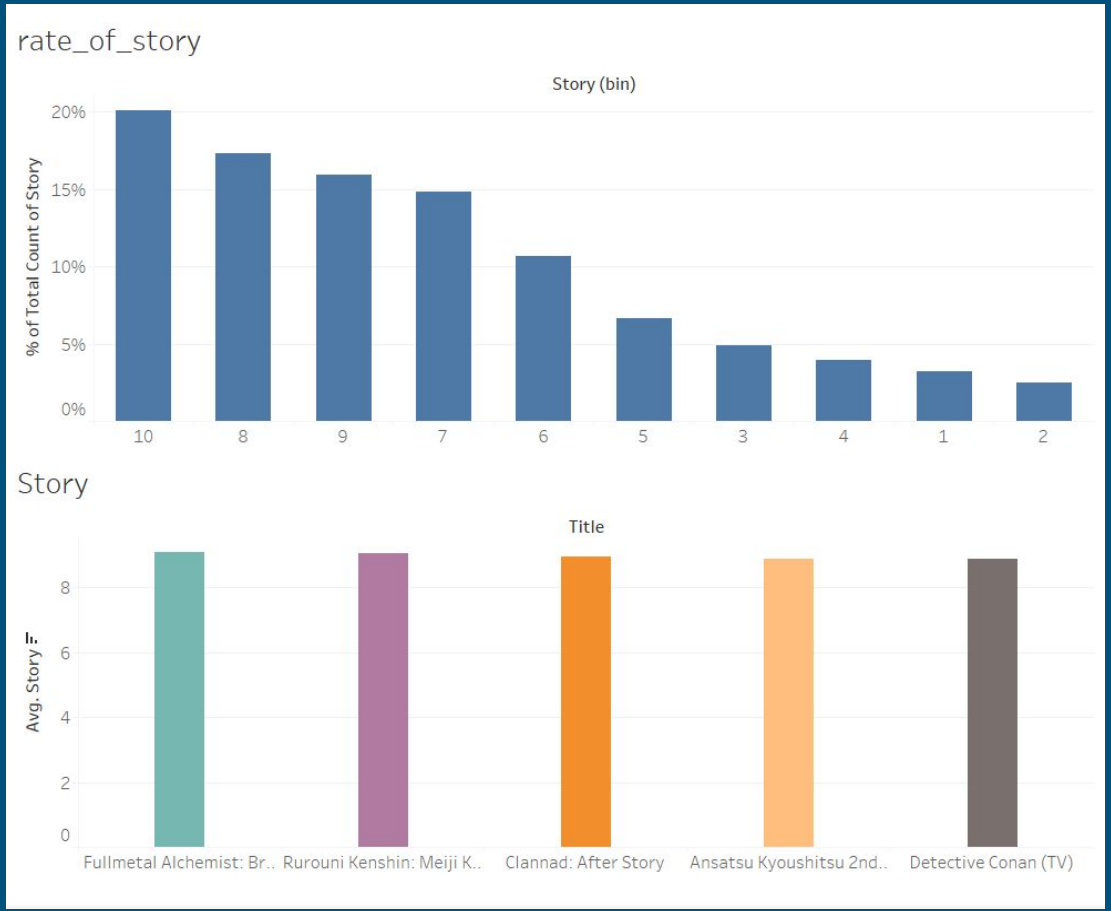
1.Animation





- Here the percentage of users who have given a rating of 10 is 23%.
- 22% have given a rating of 8.
- Anything below six should not be considered.
- The top 5 animes with high average score in the category "*animation*".
- The condition is that the animes should have been watched by at least 100 users.

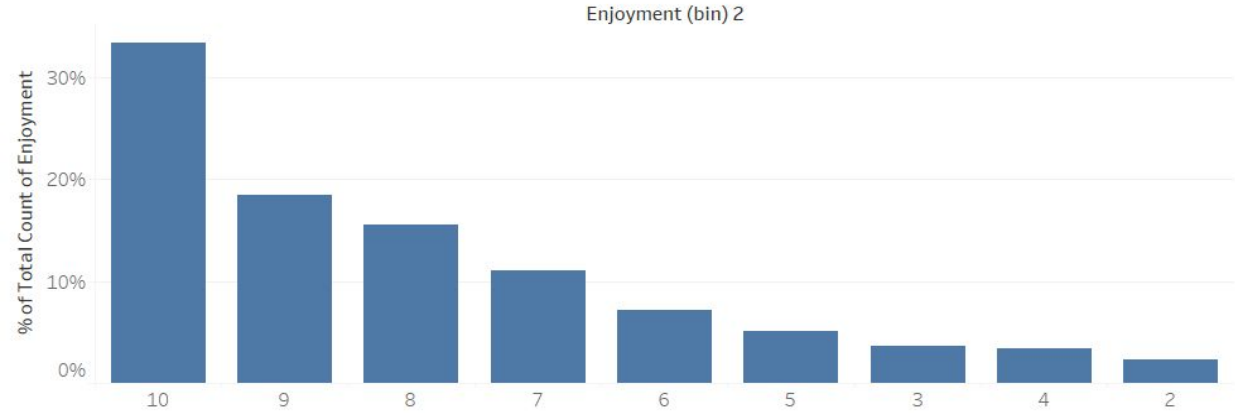
2.Story



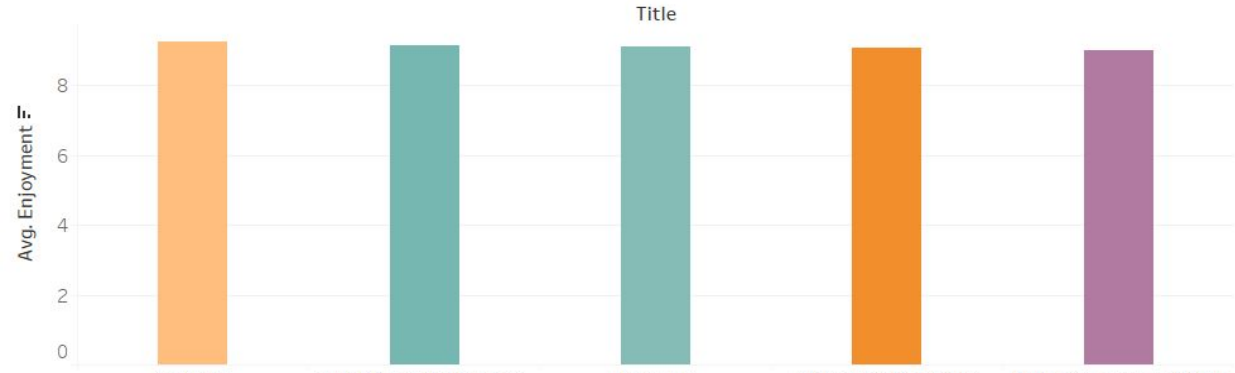
- Here the percentage of users who have given a rating of 10 is 20%.
- 17.5 % have given a rating of 8.
- Anything below six should not be considered.
- It can be observed that users care more about animation than story (look at the decreasing trend in rating)
- The top 5 animes with high average score under the category ‘***story***”.
- Here again the condition by which these animes are top 5 is that they should have been watched by at least 100 users.

3.Enjoyment

enjoyment_rate



Top_animes_enjoyment



- From the graph it can be observed that users are particular about whether they enjoy an anime or not.
- It can be seen that 35% of users have rated 10 for various anime under enjoyment.
- So anything below 6 should not be considered.
- The top 5 animes with high average score under the category ***"enjoyment"***.
- Here again the condition by which these animes are top 5 is that they should have been watched by at least 100 users.

Code

Below are some code snapshots which I used to gain insights for my analysis

```
select distinct a.title,
count(distinct r.uid) as no_of_users,
round(avg(r.story),1) as avg_story
from anime_new a
left join review r on a.anime_id=r.anime_id
group by title
having avg_story > 7 and no_of_users >100
order by avg_story DESC;
```

```
select distinct a.title,
count(distinct r.uid) as no_of_users,
round(avg(r.score),2) as avg_score
from anime_new a
left join review r on a.anime_id=r.anime_id
left join profile p on r.profile=p.profile
where gender ='Female'
group by a.title
having no_of_users >100
order by avg_score DESC
limit 10;
```

```
select distinct a.title,
count(distinct r.uid) as no_of_users,
round(avg(r.score),2) as avg_score
from anime_new a
left join review r on a.anime_id=r.anime_id
left join profile p on r.profile=p.profile
where gender not in ('Female','Male')
group by a.title
having no_of_users >100
order by avg_score DESC;
```

```
select distinct a.title,
count(distinct r.uid)as no_of_users,
round(avg(r.score),2) as avg_score
from anime_new a
left join review r on a.anime_id=r.anime_id
where genre='Comedy'
group by title
having no_of_users >100
order by avg_score DESC;
```

```
select distinct a.title,
count(distinct r.uid) as no_of_users,
round(avg(r.animation),1) as avg_animation
from anime_new a
left join review r on a.anime_id=r.anime_id
group by title
having avg_animation > 7 and no_of_users >100
order by avg_animation DESC;
```

```
select distinct a.title,
count(distinct r.uid) as no_of_users,
round(avg(r.sound),1) as avg_sound
from anime_new a
left join review r on a.anime_id=r.anime_id
group by title
having avg_sound> 7 and no_of_users >100
order by avg_sound DESC;
```

```
select distinct a.title,
count(distinct r.uid)as no_of_users ,
round(avg(r.score),2) as avg_score,
a.popularity
from anime_new a
left join review r on a.anime_id=r.anime_id
group by title,a.popularity
having no_of_users >100
order by a.popularity ;
```

Appendix

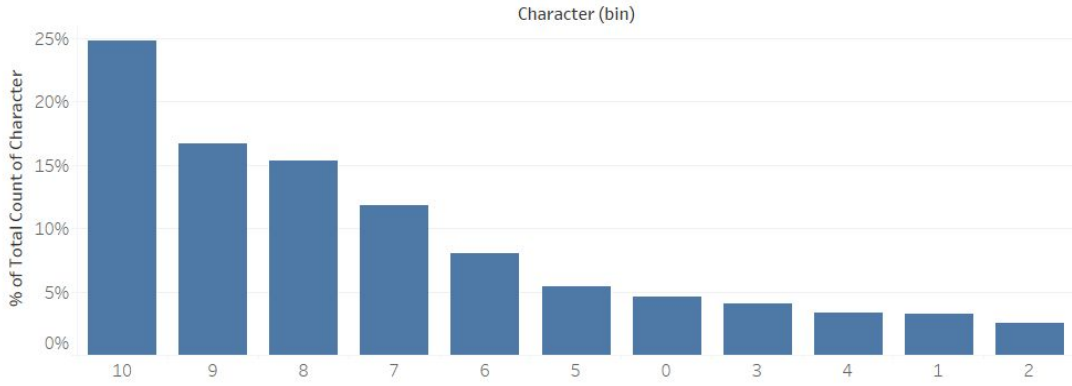
The tools used for this project are

- Microsoft excel
- Tableau
- Mysql workbench

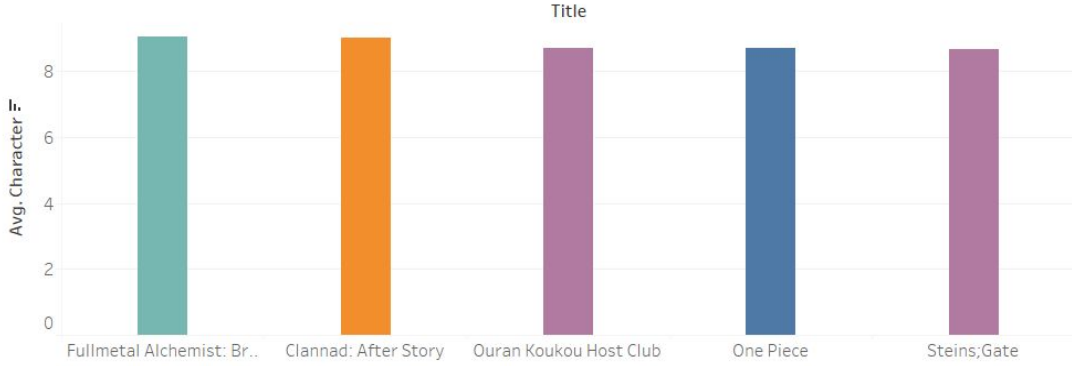
Additional information about the other categories including sound and characters are below.

4.Characters

rate_of_characters

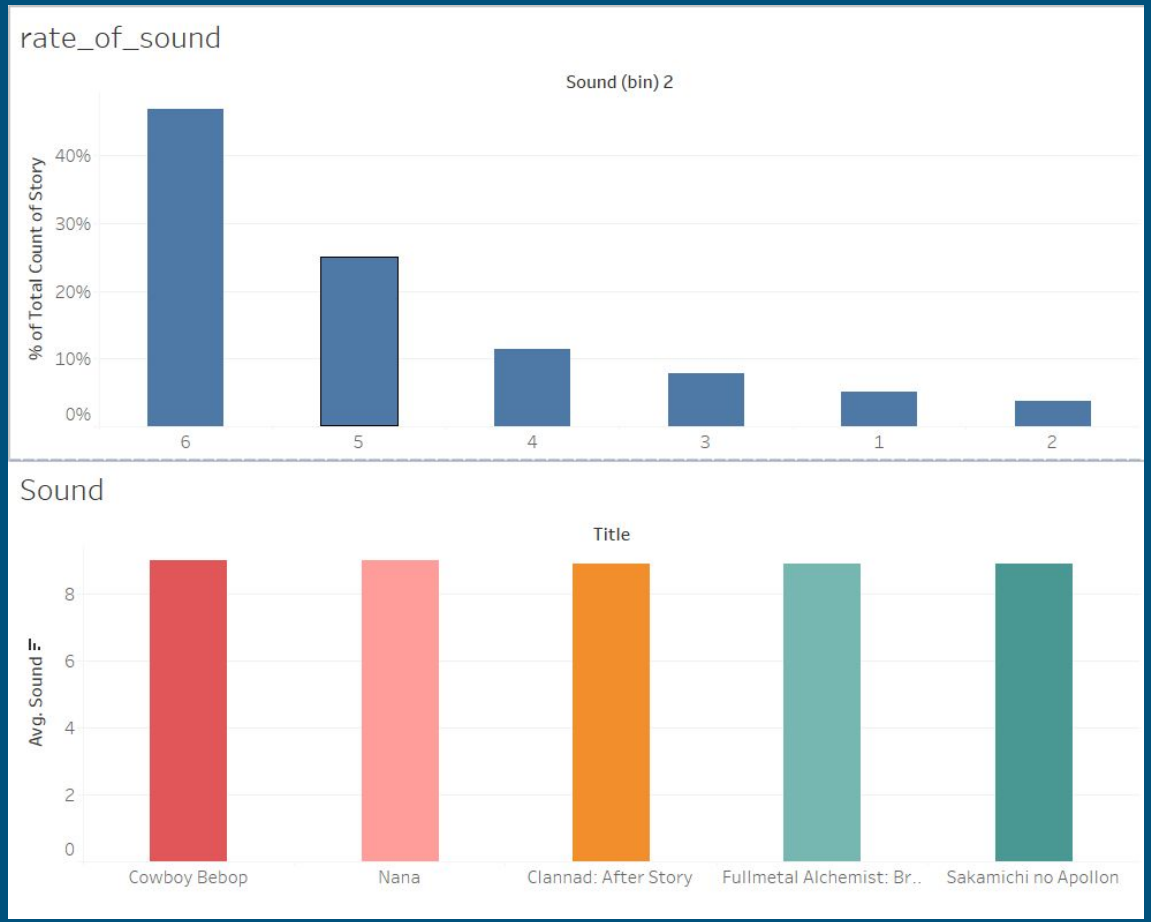


Top_anime_characters



- Here the percentage of users who have given a rating of 10 is 25%.
- 16.5 % have given a rating of 9.
- There is a drastic drop in the ratings
- Anything below six should not be considered.
- The top 5 anime with high avg score in "*characters*".
- Here again the condition by which these animes are top 5 is that they should have been watched by at least 100 users.

5.Sound



- Here the percentage of users who have given a rating of 6 is 45%
- Then 24.9% have given a rating of 8
- It can be observed that users are not really liking the sound quality of the anime
- Either the voices quality is not good or the music of the anime.This needs further analysis
- The top 5 anime with high avg score in "*sound*"
- Here again the condition by which these animes are top 5 is that they should have been watched by at least 100 users.

Questions?