“AnimeLens: A Hybrid Approach for Personalized Anime Recommendations using Deep Collaborative Filtering and TF-IDF Content Based Filtering.”

Submitted by

Group Name: **Troopers**

Submitted to

Farhana Akter Bina

Assistant Professor

Submission Date: May 05, 2024

Department of Statistics & Data Science

Jahangirnagar University

Table of Contents

1. Introduction

In today's digital age, where the abundance of information overwhelms users, recommendation systems play a pivotal role in assisting individuals in discovering relevant content, products, or services (Ekstrand, 2011). These systems leverage data-driven algorithms to predict and suggest items that users are likely to find appealing or useful, thereby enhancing user experience, engagement, and satisfaction across various online platforms. The exponential growth of e-commerce, social media, streaming services, and other online platforms has intensified the need for effective recommendation systems. From personalized product recommendations on e-commerce websites to tailored content suggestions on streaming platforms, recommendation systems have become indispensable tools for businesses aiming to enhance user engagement, drive sales, and foster customer loyalty. Moreover, recommendation systems operate within ethical and privacy frameworks, necessitating careful consideration of user privacy, transparency, and fairness. As recommendation systems exert significant influence on user behavior and decision-making processes, ensuring ethical and responsible design principles is paramount to maintaining user trust and societal well-being (Ekstrand, 2011).

In the realm of entertainment, the world of anime presents a vast and diverse landscape, offering a multitude of genres, themes, and storytelling styles to suit every viewer's taste. However, navigating this extensive catalog of anime series and films can be daunting, particularly for newcomers or those seeking personalized recommendations tailored to their preferences. In response to this challenge, anime recommendation systems emerge as invaluable tools, leveraging data-driven algorithms to assist users in discovering anime titles that resonate with their interests, preferences, and viewing habits. The popularity of anime continues to soar globally, attracting audiences of all ages and cultural backgrounds. From classic shonen adventures to slice-of-life dramas and fantastical epics, the appeal of anime transcends geographical boundaries, making it a thriving and dynamic entertainment medium. As the demand for anime content grows, so does the need for effective recommendation systems that can guide users through the vast array of available titles, helping them discover hidden gems and new favorites along the way (Ko et al., 2022).

The landscape of recommendation systems encompasses various algorithms and methodologies, including collaborative filtering, content-based filtering, matrix factorization, deep learning, and hybrid approaches. Each approach has its strengths and weaknesses, making it essential for researchers and practitioners to explore and evaluate different techniques to determine the most suitable solution (Singhal et al., 2017). The core objective of our project is to develop and evaluate an anime recommendation system tailored specifically to the unique characteristics of the anime fandom. By harnessing the power of machine learning algorithms, collaborative filtering techniques, and user engagement data, our recommendation system aims to provide anime enthusiasts with personalized recommendations that reflect their individual tastes, preferences, and viewing history (Kumar & Thakur, 2018). Through this project, we seek to enhance the anime viewing experience, promote content discovery, and foster a deeper connection between users and the rich tapestry of anime storytelling.

Our objective is to advance the field of anime recommendation systems by offering valuable insights and a practical framework for researchers, developers, and enthusiasts. We aim to empower anime fans with an effective tool that enhances their viewing experience and enables exploration within the diverse world of anime entertainment. Leveraging a range of algorithms and methodologies, including user-based filtering, collaborative filtering, content-based filtering, matrix factorization, deep learning, and hybrid approaches, our goal is to identify and implement the most suitable solutions for personalized anime recommendations.

2. Methodology

2.1 Data Collection

The dataset utilized for this project was obtained from Kaggle, a prominent platform for data science and machine learning resources. The dataset contains comprehensive information about anime titles, user details, and user scores, enabling in-depth analysis and insights into anime recommendations and user preferences. The dataset was collected from the following link: MyAnimeList Dataset on Kaggle (Anime Dataset 2023, 2023)**).**

The dataset contains mainly three files:

- **Anime Dataset:** The dataset provides invaluable insights for analyzing and understanding the attributes, ratings, popularity, and audience engagement of diverse anime productions.

Attribute	Description
anime_id	Unique ID for each anime.
Name	The name of the anime in its original language.
English name	The English name of the anime.
Other name	Native name or title of the anime (can be in Japanese, Chinese, or Korean).
Score	The score or rating given to the anime.
Genres	The genres of the anime, separated by commas.
Synopsis	A brief description or summary of the anime's plot.
Type	The type of the anime (e.g., TV series, movie, OVA, etc.).
Episodes	The number of episodes in the anime.
Aired	The dates when the anime was aired.
Premiered	The season and year when the anime premiered.
Status	The status of the anime (e.g., Finished Airing, Currently Airing, etc.).
Producers	The production companies or producers of the anime.
Licensors	The licensors of the anime (e.g., streaming platforms).
Studios	The animation studios that worked on the anime.
Source	The source material of the anime (e.g., manga, light novel, original).
Duration	The duration of each episode.
Rating	The age rating of the anime.
Rank	The rank of the anime based on popularity or other criteria.
Popularity	The popularity rank of the anime.
Favorites	The number of times the anime was marked as a favorite by users.
Scored By	The number of users who scored the anime.
Members	The number of members who have added the anime to their list on the platform.
Image URL	The URL of the anime's image or poster.

- **User Details:** The User Details Dataset offers valuable insights into user behavior and preferences on the anime platform, including mean scores, genre preferences, and user segmentation based on watching behavior.

Attribute	Description
Mal ID	Unique ID for each user.
Username	The username of the user.
Gender	The gender of the user.
Birthday	The birthday of the user (in ISO format).
Location	The location or country of the user.
Joined	The date when the user joined the platform (in ISO format).
Days Watched	The total number of days the user has spent watching anime.
Mean Score	The average score given by the user to the anime they have watched.
Watching	The number of anime currently being watched by the user.
Completed	The number of anime completed by the user.
On Hold	The amount of anime on hold by the user.
Dropped	The number of anime dropped by the user.
Plan to Watch	The number of anime the user plans to watch in the future.
Total Entries	The total number of anime entries in the user's list.
Rewatched	The number of anime rewatched by the user.
Episodes Watched	The total number of episodes watched by the user.

- **User Score:** Various analyses and insights into user interactions with anime are enabled by the User Score Dataset. Highly rated and popular anime among users can be identified by examining user ratings for different anime titles.

Attribute	Description
user_id	Unique ID for each user.
Username	The username of the user.
anime_id	Unique ID for each anime.
Anime Title	The title of the anime.
rating	The rating given by the user to the anime.

The information contained in this dataset is extremely helpful in gaining a knowledge of the preferences and levels of interaction of users on the anime platform.

2.2 Data Pre-Processing

The Min-Max scaler transforms the rating column to a range between 0 and 1, inclusive. This preprocessing step is significant as it standardizes the rating values, ensuring uniformity and comparability across the dataset. The equation for Min-Max scaling in passive form is expressed as:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Here,

X_{scaled} = represents the scaled rating value
X = denotes the original rating value
X_{min} = minimum rating value in the dataset
X_{max} = maximum rating value in the dataset

Min-Max scaling preserves the relative differences between ratings while accommodating them within a consistent range, facilitating accurate analysis and modeling without undue influence from the original scale.

2.3 Exploratory Data Analysis

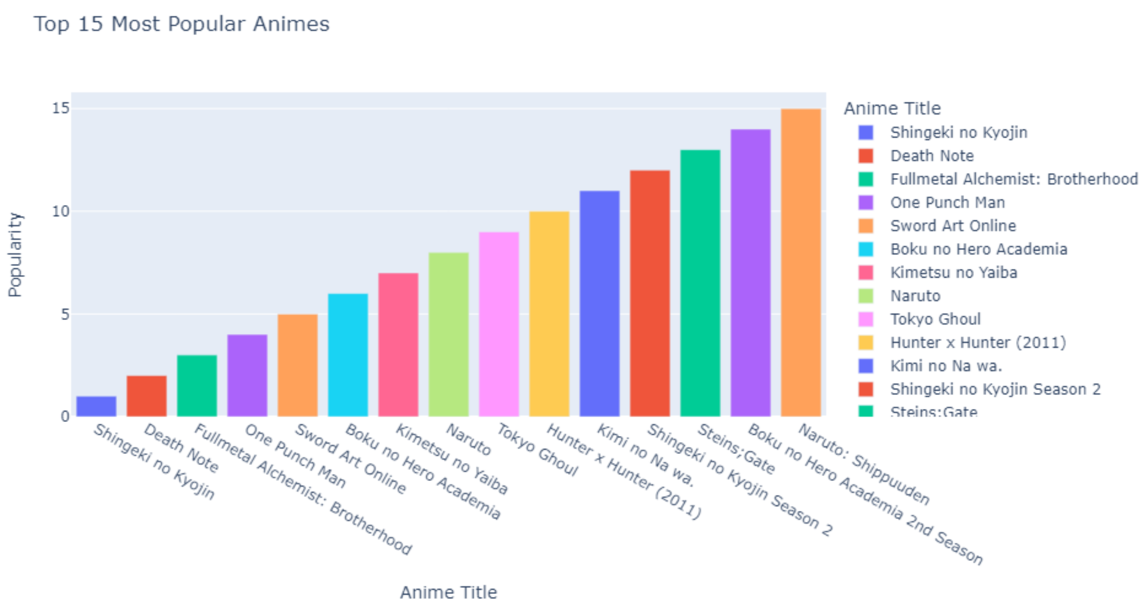

Figure 01: Bar plot of the Top 15 Most Popular Anime Series.

The bar graph highlights the most popular anime series, with "Shingeki no Kyojin Season 2" and "Steins;Gate" leading the pack as the top favorites. On the other end, classics like "Shingeki no Kyojin" and "Death Note" also make the list but with relatively lower popularity scores.

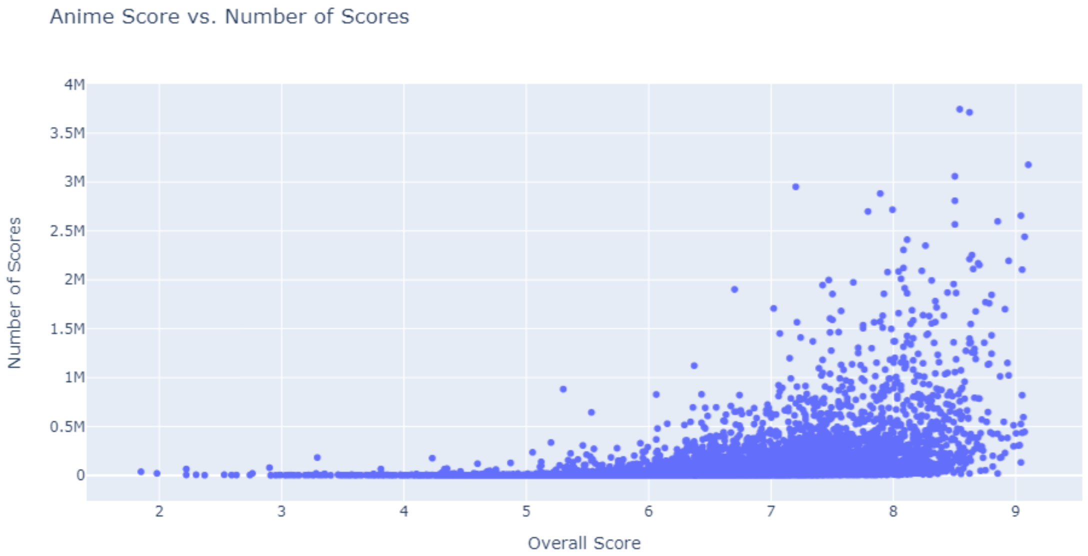

Figure 02: Scatter plot of overall Score.

The scatter plot illustrates the relationship between anime scores and the number of scores received. It shows that most anime is rated between 6 and 8, with a significant number of ratings clustering around higher scoring animes, indicating a trend where more popular anime tend to receive higher scores.

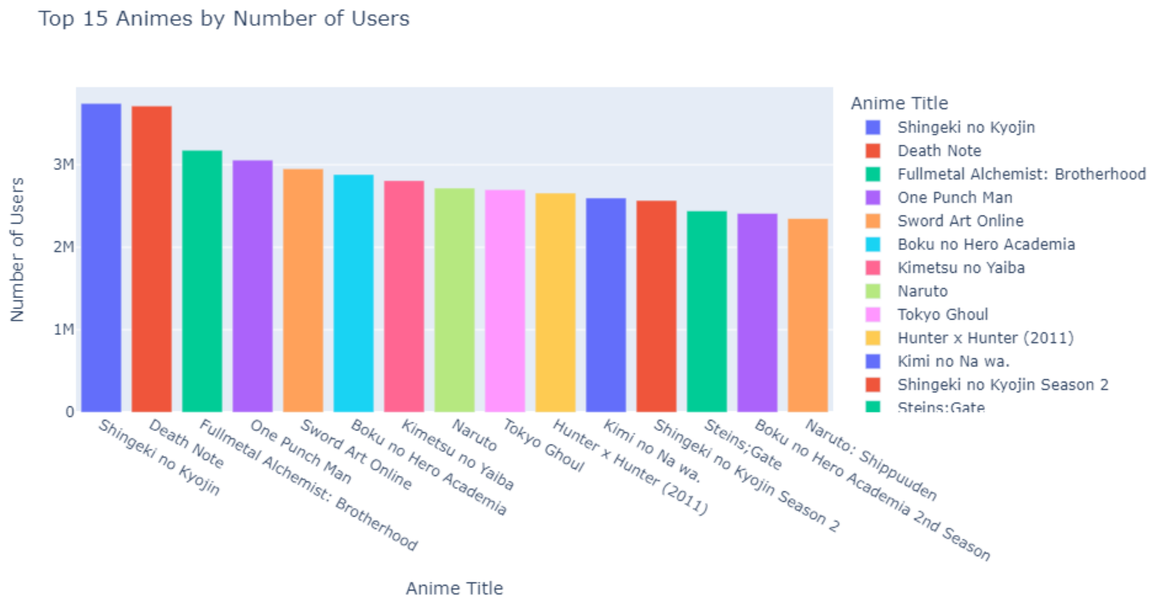

Figure 03: Bar plot of Number of user and Anime title.

The bar graph showcases the top 15 anime series by user engagement, revealing that "Shingeki no Kyojin" and "Death Note" lead with over 3 million users each. The graph displays a diverse range of popular anime, with other notable entries like "Fullmetal Alchemist: Brotherhood" and "One Punch Man" closely following in terms of user numbers.

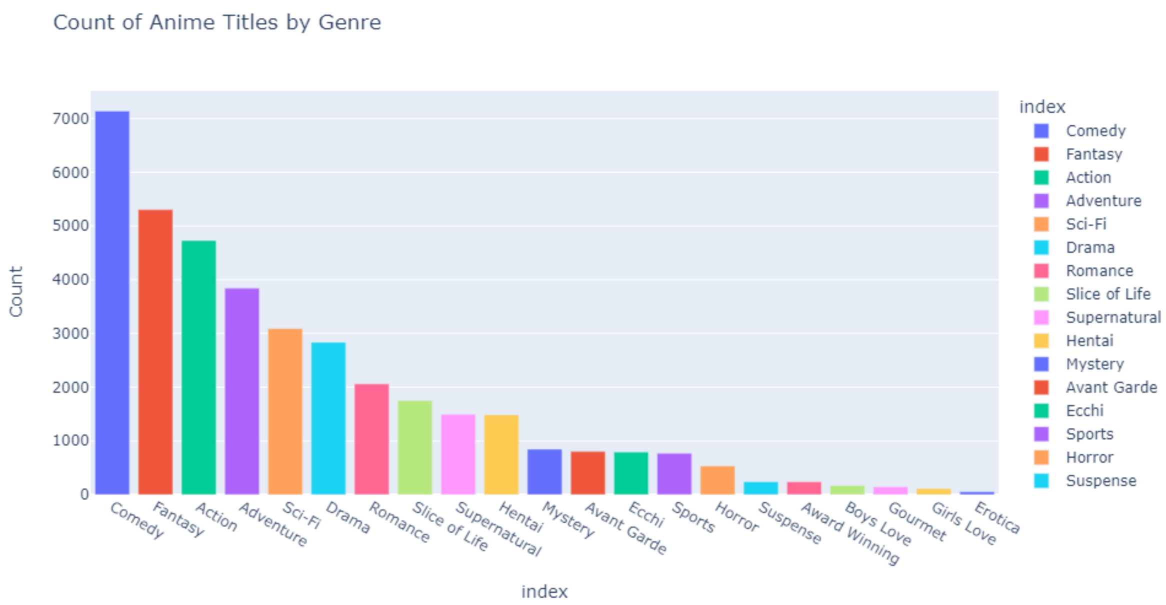

Figure 04: Bar plot of the Anime title by genre.

The bar graph displays the distribution of anime titles across various genres. Comedy is the most populated genre, followed by Fantasy and Action, each with several thousand titles. The graph shows a broad diversity in genre popularity, with niches like Horror, Suspense, and more specialized categories like Avant Garde and Erotica having significantly fewer titles.

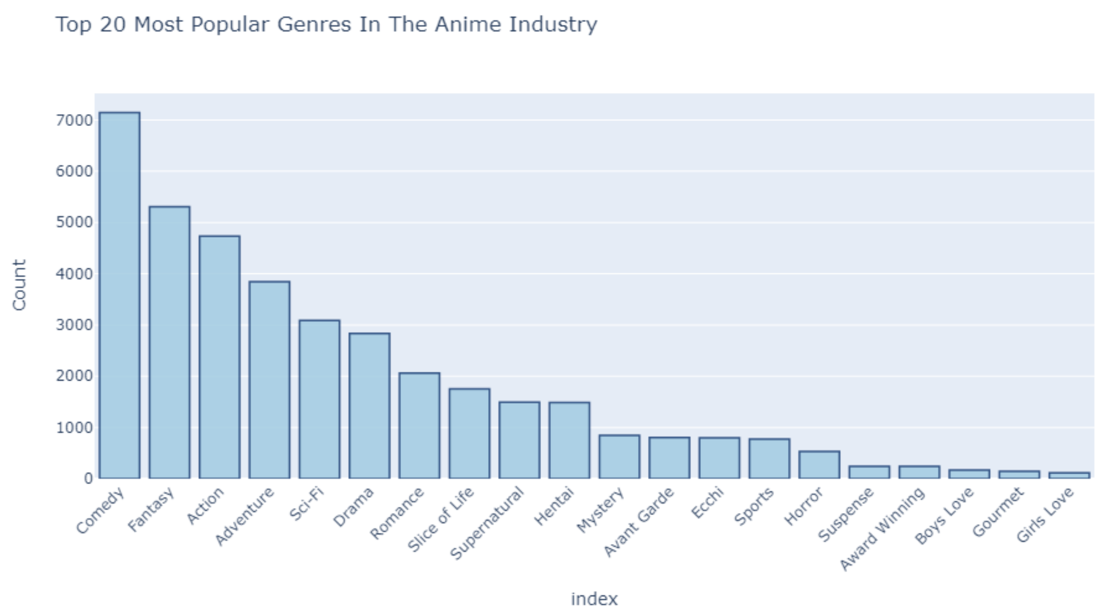

Figure 05: Bar plot of the popular genres in the anime industry.

The bar graph provides an insightful look at the top 20 most popular anime genres, with Comedy leading by a substantial margin. The following closely are Fantasy, Action, and Adventure, showing strong preferences among anime viewers. The genres display a clear gradient in popularity, with more niche categories like Sports, Horror, and various specialized themes like "Boys' Love" and "Girls' Love" appearing less frequently but still ranking among the top genres in the industry.

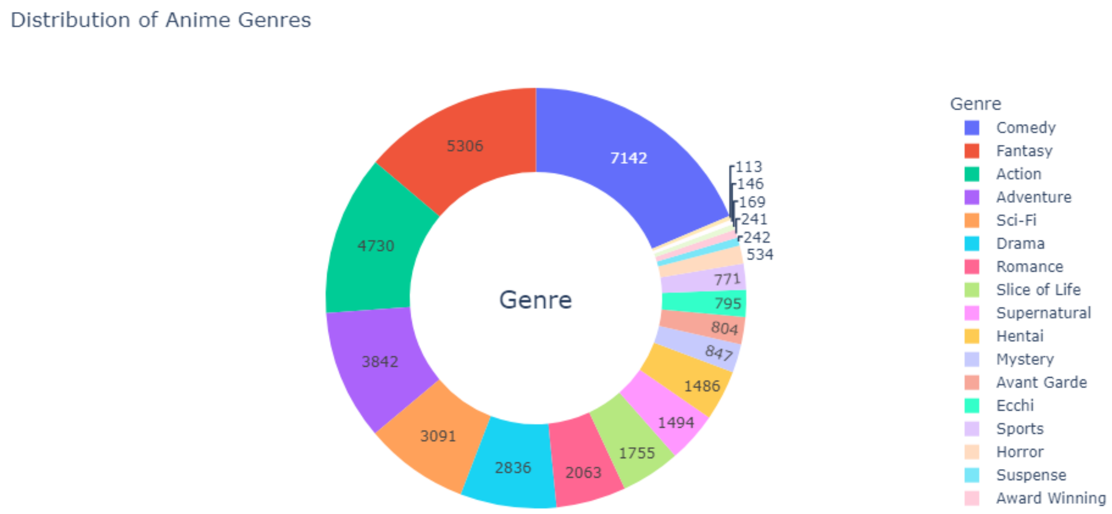

Figure 06: Pie chart of distribution of anime genres.

The pie chart vividly illustrates the distribution of anime genres. Comedy emerges as the most represented genre with 7142 titles, significantly dominating the anime landscape. Fantasy and Action also have strong showings, each comprising a large segment of the circle, while other genres like Romance, Sci-Fi, and Drama maintain substantial but smaller proportions. This visual demonstrates a diverse range of anime offerings, catering to varied audience preferences.

Figure 07: Word embedding plot of genres.

The word embedding plot visually maps the relationship and prominence of genres in the anime industry. Large, central terms like "Comedy," "Fantasy," and "Adventure" highlight their dominance. The proximity of certain genres, such as "Action" near "Adventure" and "Comedy" near "Romance," suggests thematic similarities, providing insights into how genres overlap and interact.

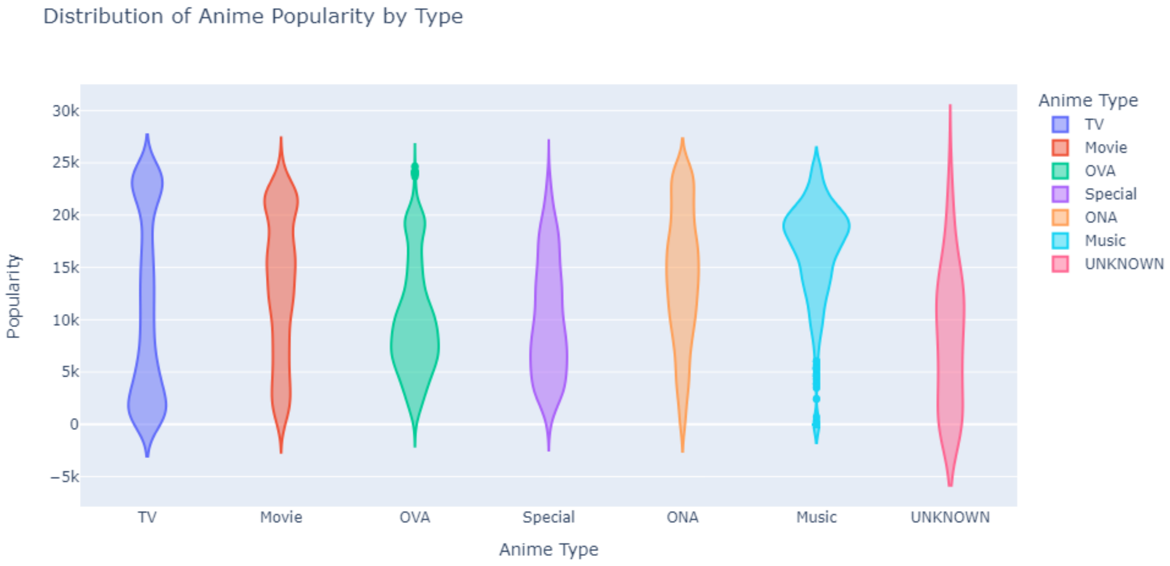

Figure 08: Violine plot of the distribution of anime popularity by types.

The violin plot illustrates the distribution of popularity for various anime types, showing that TV shows and movies have the widest range of popularity, with significant peaks. In contrast, types like OVA, Special, and ONA display more uniform popularity levels. The "Music" and "UNKNOWN" categories show narrow distributions, indicating less variation in their popularity.

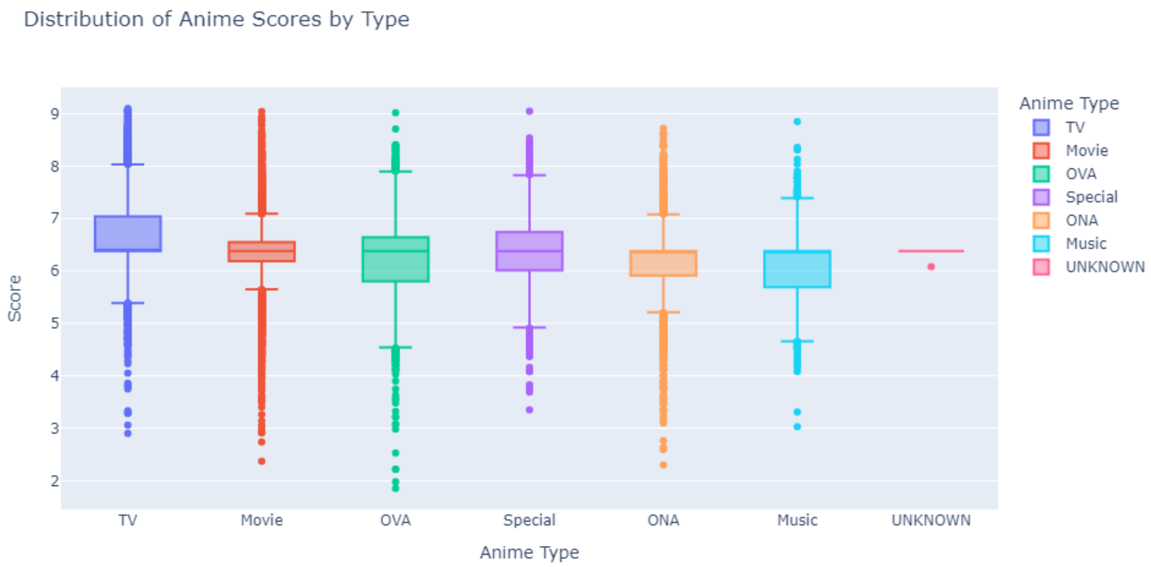

Figure 09: Box plot of the distribution of anime scores by types.

The box plot illustrates how scores are distributed across different anime types, showing that TV series and movies generally receive higher median scores around 7, with a broader spread for TV series. Other formats like OVAs, Specials, and ONAs also perform well, albeit with slightly lower medians, while Music and the UNKNOWN category exhibit more consistency but narrower score ranges.

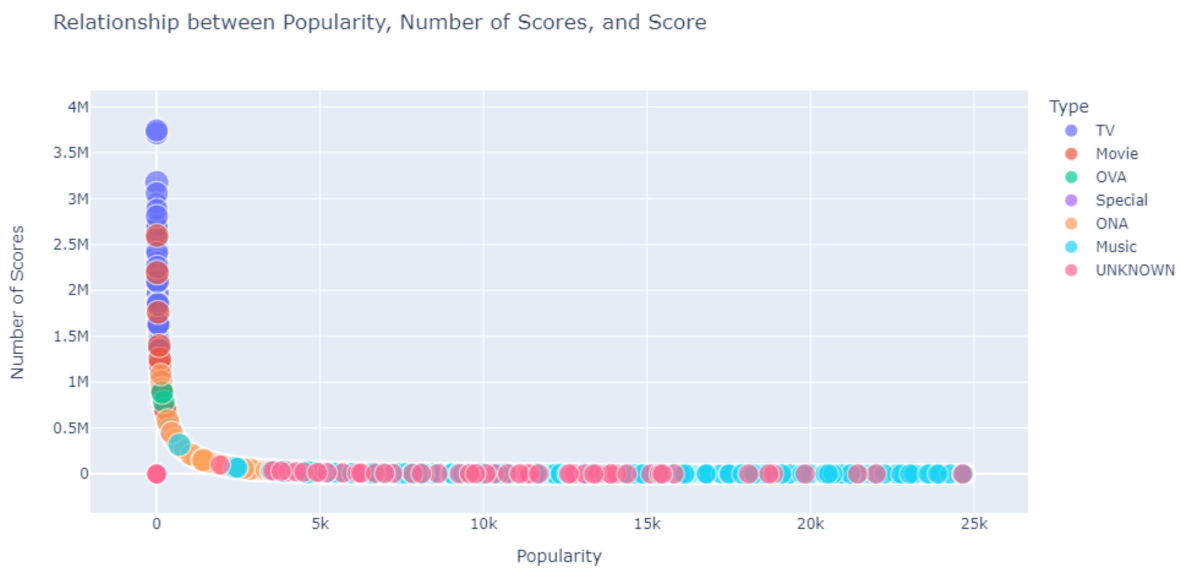

Figure 10: Bubble plot of popularity, number of scores and scores.

The bubble chart shows that most anime, especially TV series, accumulate many scores but lower popularity ratings. This visualization highlights the challenge for anime titles to achieve both high popularity and extensive viewer ratings.

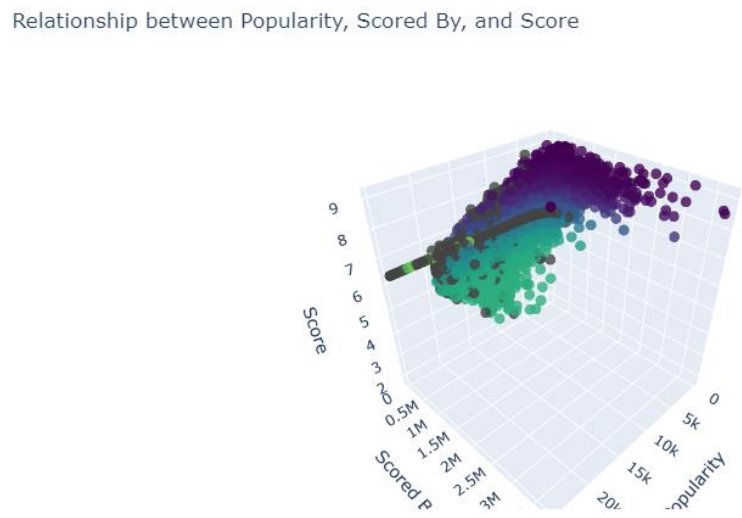

Figure 11: 3D scatter plot of popularity, scored by and score.

The 3D scatter plot shows a dense clustering of anime titles, demonstrating that most titles rated by a moderate to high number of users tend to have scores between 6 and 8. As popularity decreases, the variability in scores increases, indicating a wider range of ratings for less popular anime. This visualization effectively captures the intricate dynamics between user engagement, popularity, and the perceived quality of anime.

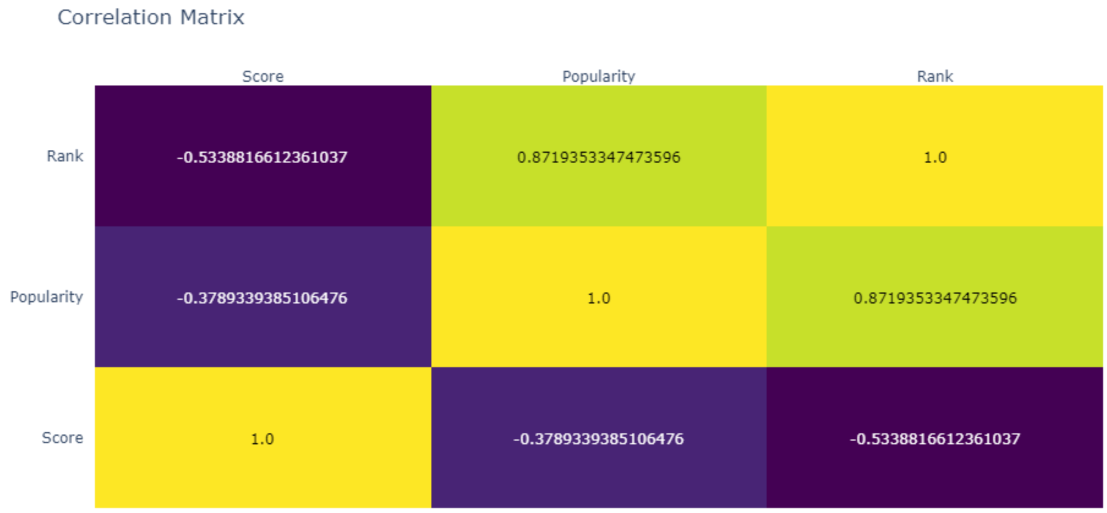

Figure 12: Correlation plot of the score, popularity, and rank.

The correlation matrix visually represents the relationships between score, popularity, and rank of anime. It indicates a strong negative correlation (-0.54) between rank and score, suggesting that higher scores generally correspond to better (lower) ranks. Similarly, there's a strong positive correlation (0.87) between rank and popularity, showing that more popular anime tend to have better ranks. The relationship between score and popularity is weaker and negative (-0.38), indicating that higher scores do not always align with higher popularity.

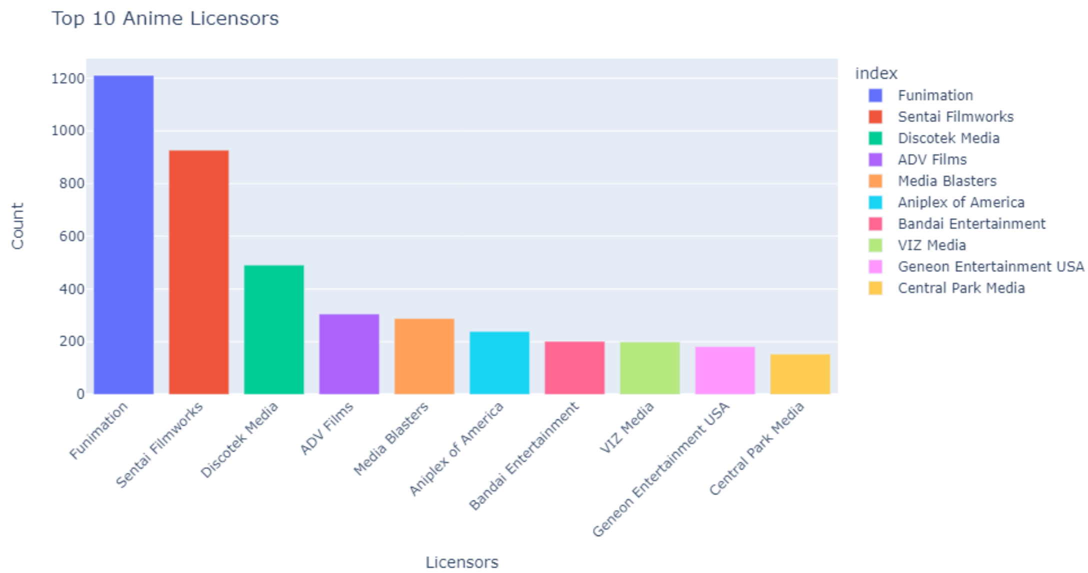

Figure 13: Bar plot of the top anime licensors.

The bar graph illustrates the dominance of Funimation in the anime licensing industry, licensing over 1000 titles, far surpassing its nearest competitors, Sentai Filmworks and Discotek Media. This visualization highlights the varying scale of operations among the top 10 anime licensors.

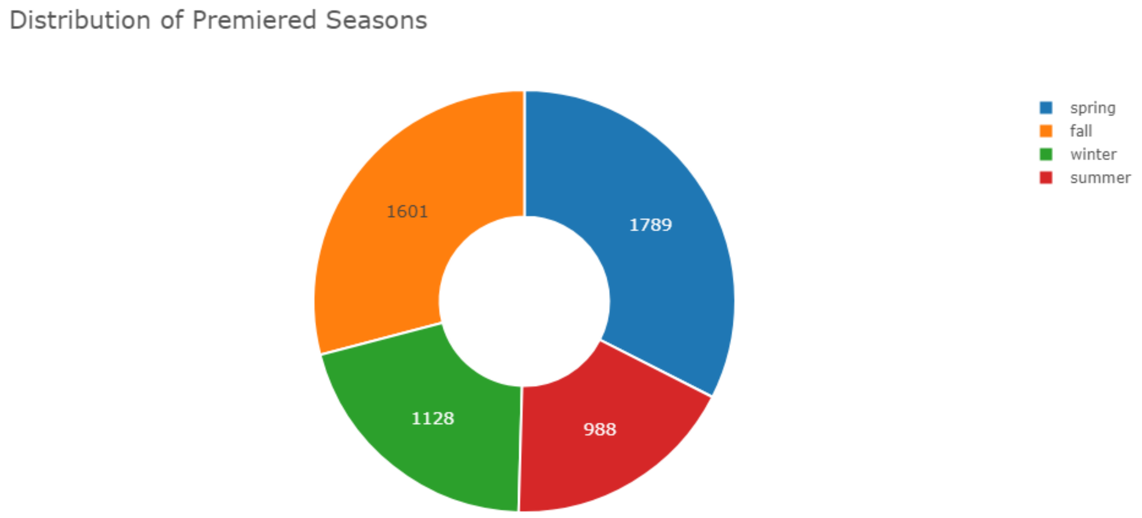

Figure 14: Pie chart of the distribution of premiered seasons.

The pie chart illustrates the distribution of anime premieres across different seasons. Spring sees the highest number of premieres with 1789 shows, followed by Fall with 1601 shows. Winter and Summer have fewer premieres, with 1128 and 988 shows respectively, indicating a trend where more new anime titles are released during the spring and fall seasons.

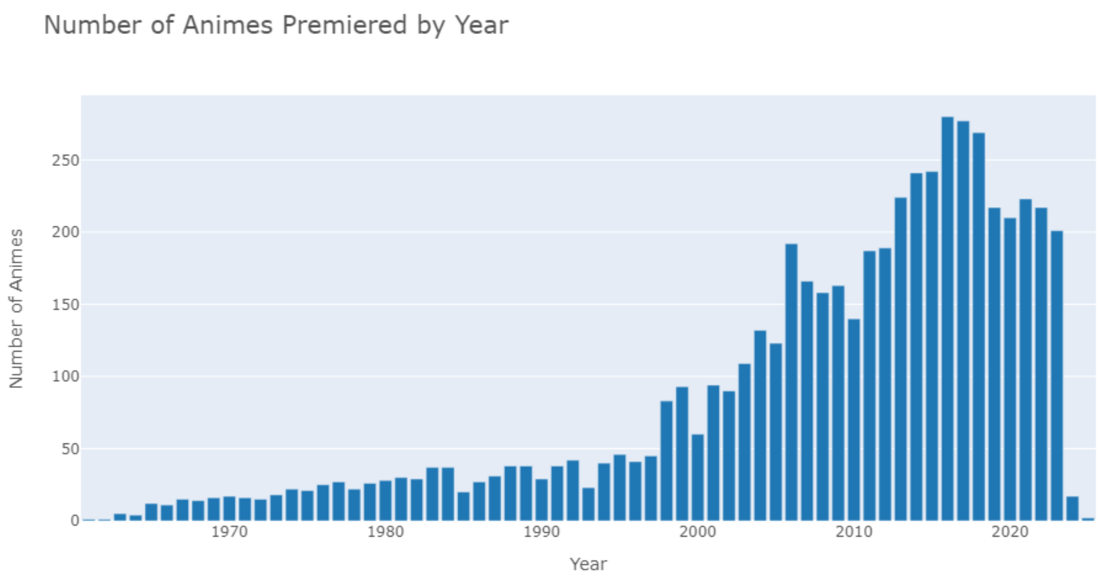

Figure 15: Bar plot number of animes premiered by year.

The bar graph depicts a significant rise in anime premieres starting in the early 2000s, peaking around 2015, and followed by a gradual decline. This trend underscores the expansion and subsequent stabilization of the anime industry over recent decades.

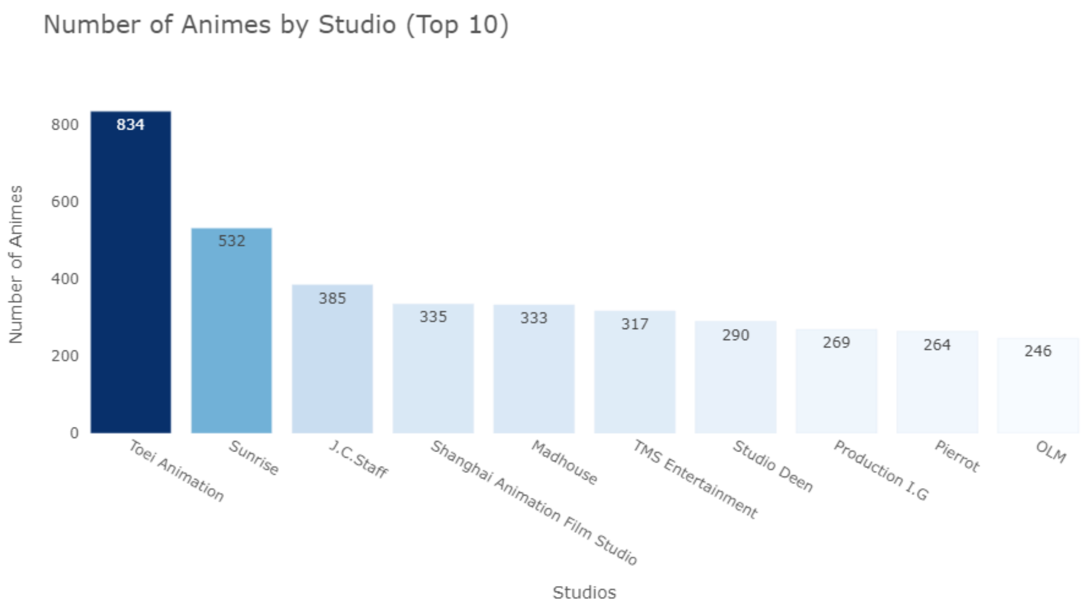

Figure 16: Bar plot of the number of animes by studio.

The bar graph displays the production output of the top 10 anime studios, with Toei Animation leading significantly by producing 834 anime titles. Sunrise and J.C.Staff follow with 532 and 385 titles, respectively. The chart shows a steep decline in output as the list progresses, highlighting the dominance of a few key players in the anime production industry.

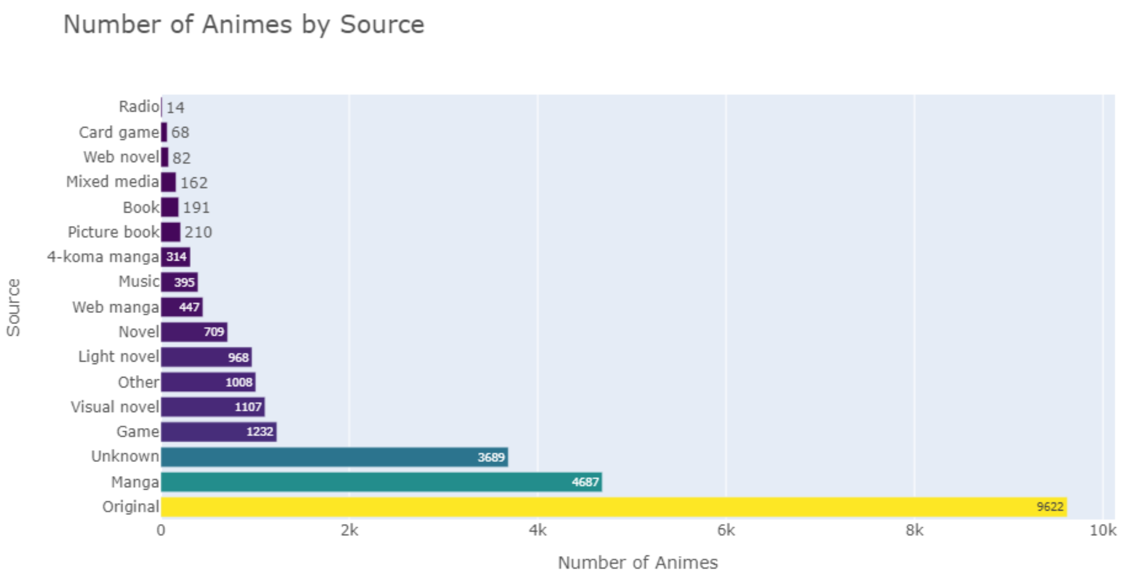

Figure 17: Bar plot of the number of animes by source.

The bar graph highlights that the majority of anime are based on original concepts, with 9622 titles, followed by manga adaptations totaling 4687. The data demonstrates the industry's reliance on a variety of sources, including visual novels, games, and light novels for content creation.

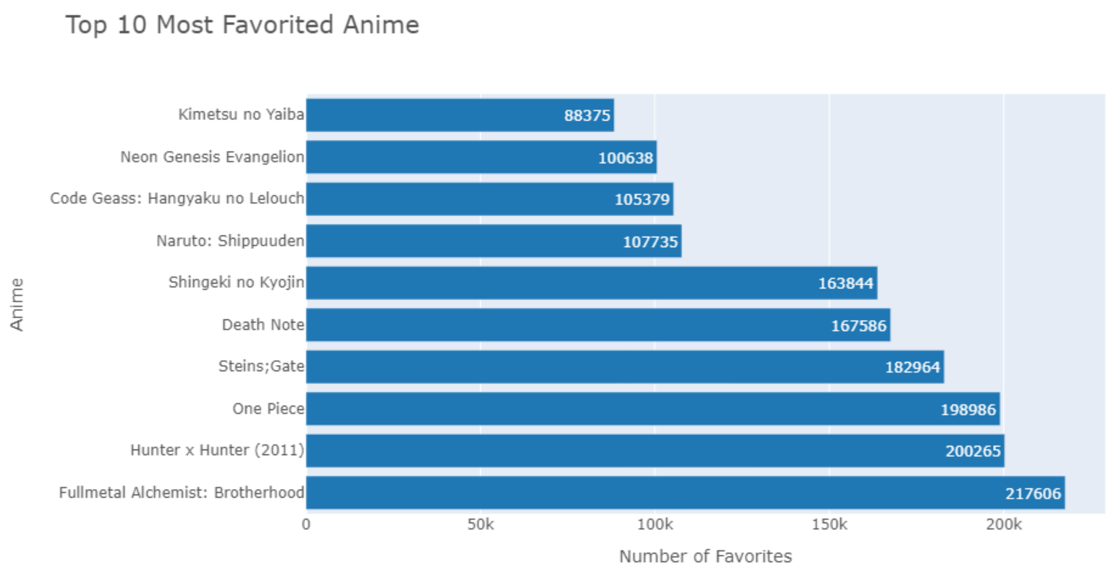

Figure 18: Bar plot of the most favorite anime.

The bar graph showcases the top 10 most favorited anime, with "Fullmetal Alchemist: Brotherhood" leading significantly, having been marked as a favorite over 217,000 times. Following closely are "Hunter x Hunter (2011)" and "One Piece," each also highly favored by viewers. This graph highlights the anime titles that have resonated most strongly with fans, gaining substantial favoritism in the community.

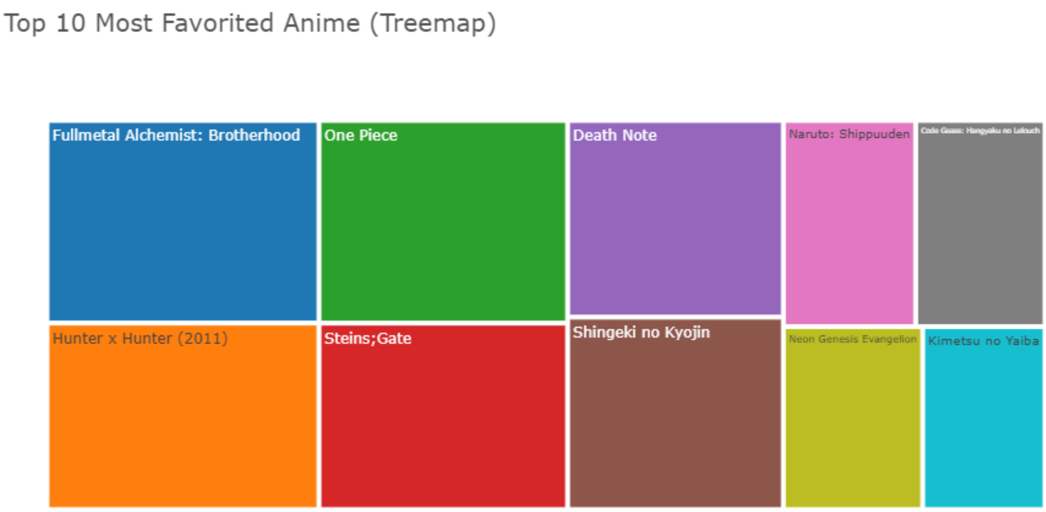

Figure 19: Tree map of the most favorite anime.

The treemap visualizes the top 10 most favorited anime, effectively using color-coded blocks to represent each series' popularity relative to each other. "Fullmetal Alchemist: Brotherhood" and "One Piece" occupy the largest areas, indicating their status as the most favorited among fans. The size of each block correlates with the number of times each title has been marked as a favorite, showcasing the significant appeal of these series within the anime community.

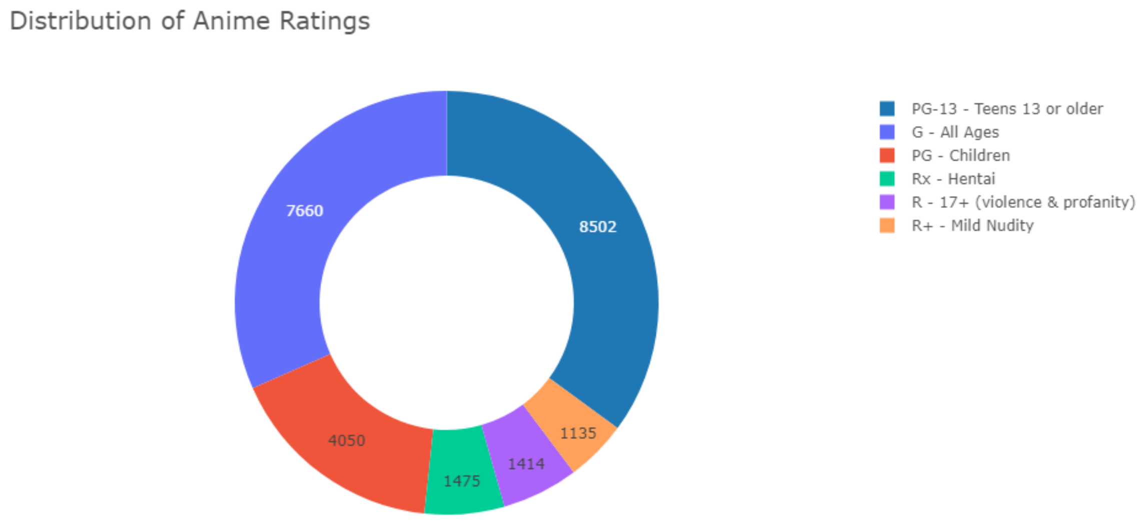

Figure 20: Pie chart of the distribution of anime ratings.

The pie chart displays the distribution of anime ratings, with "PG-13 - Teens 13 or older" being the most common, representing 8502 titles. This is followed by "G - All Ages" and "R - 17+" categories, which account for 7660 and 4050 titles respectively. The chart illustrates the range of content suitability in anime, from general audience material to more mature themes.

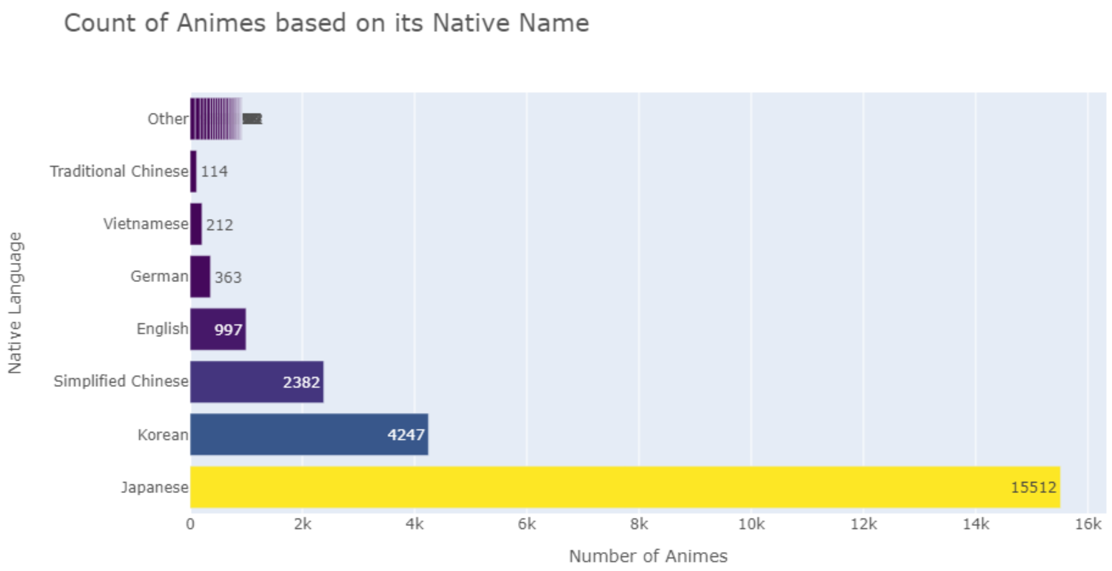

Figure 21: Bar plot of the count of anime based on its native name.

The bar graph shows that the majority of anime titles are native to the Japanese language, totaling 15,512, reflecting the genre's origins in Japan. Korean and Simplified Chinese languages also contribute notably to the anime industry but at much lower numbers.

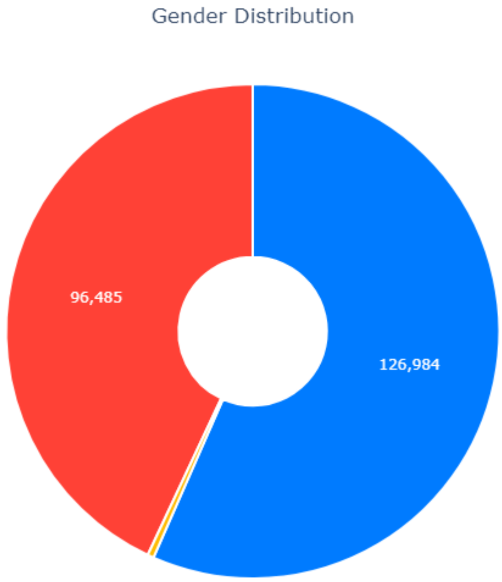

Figure 22: Pie chart of the gender distribution of the anime viewers.

The donut chart presents the gender distribution of anime viewers or characters, with the blue segment representing 126,984 and the red segment accounting for 96,485. The blue segment is larger, indicating a higher proportion of one gender over the other in the sampled population.

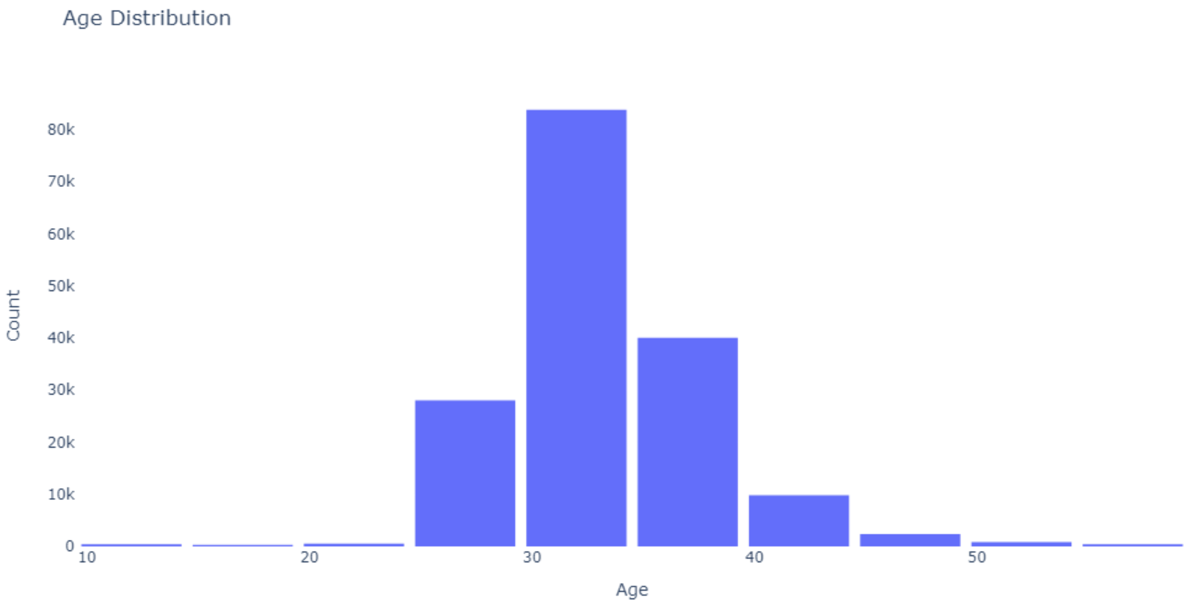

Figure 23: Bar chart of the age distribution of anime viewers.

The bar chart illustrates the age distribution of anime viewers, showing a prominent peak at the 30-year age group with the highest viewer count, and a noticeable decrease in viewership as the age increases. This highlights anime's strong appeal primarily to younger adults.

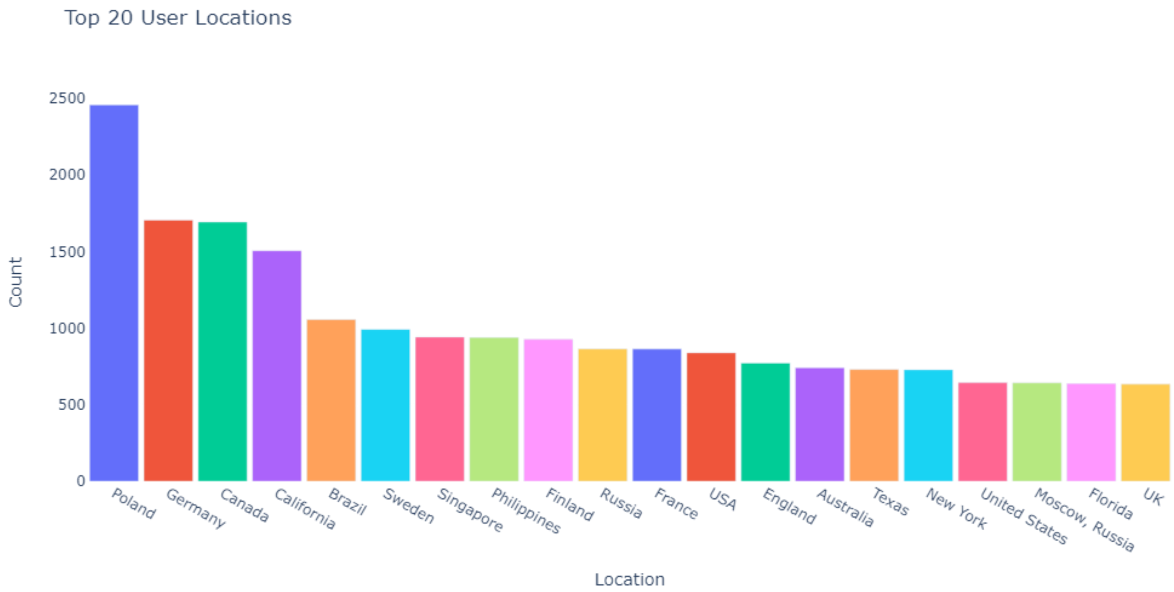

Figure 24: Bar plot of user location.

This bar graph displays the top 20 user locations by count, with Poland having the highest count, significantly more than other locations. Germany, Canada, and Brazil follow, each with slightly lower counts but comparable to one another.

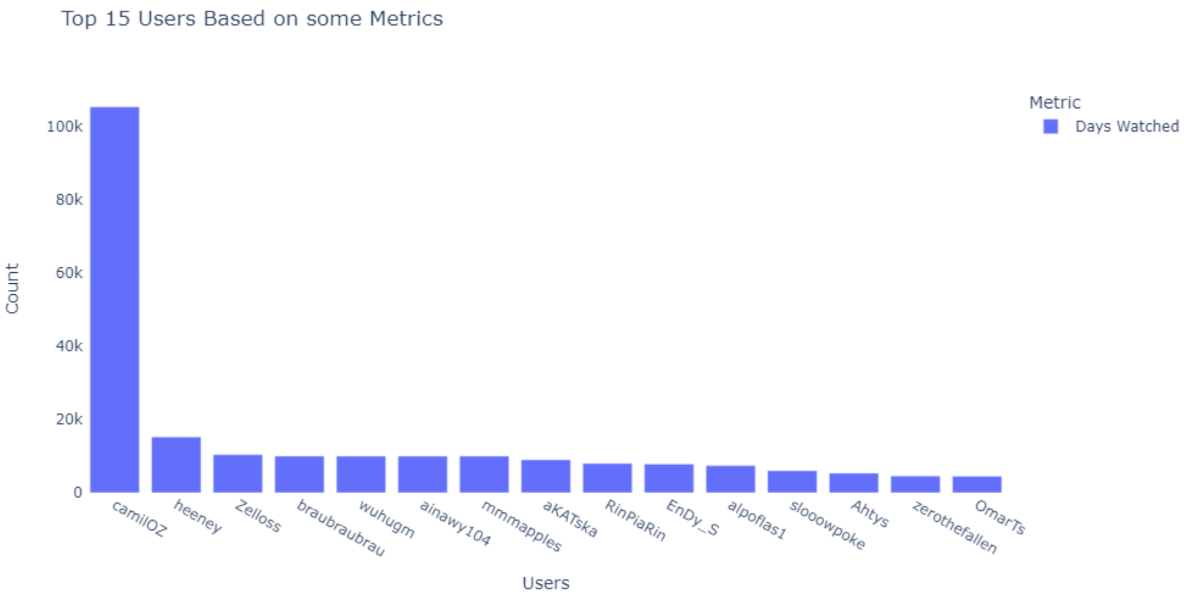

Figure 25: Bar plot of the user by days watched.

The bar graph shows the top 15 users based on the metric "Days Watched", with the user "Cammilos" having the highest count, significantly ahead of others. The subsequent users, like

"Hopeney" and "Zelkova", have considerably lower counts, and the values decrease progressively as we move towards the right of the graph.

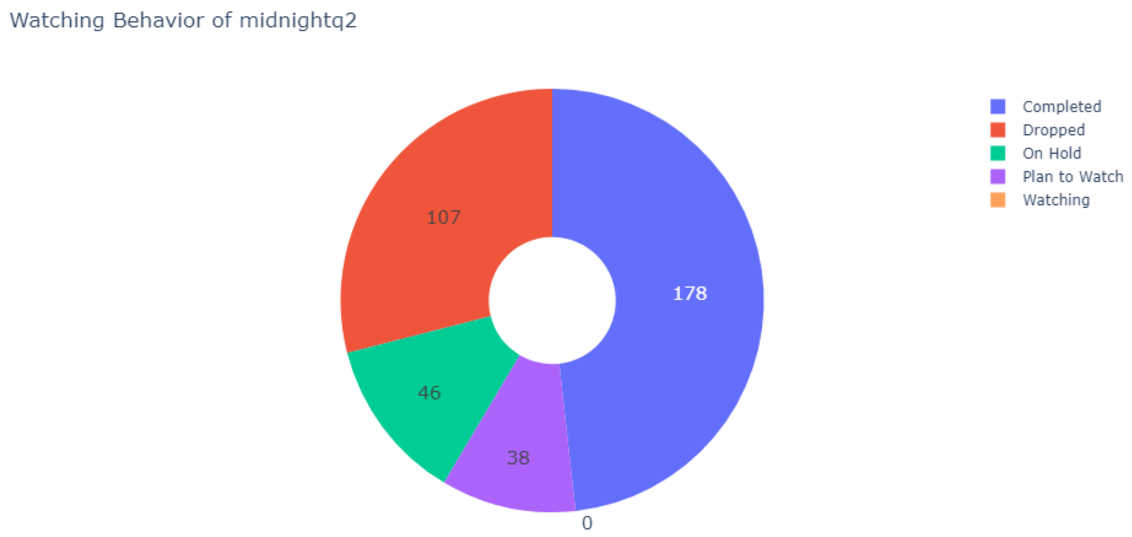

Figure 26: Pie chart of the watching behavior of the users.

The donut chart details the viewing habits of user "midnightq2," showing 178 titles completed and 107 planned for future viewing. Lesser amounts are seen in "Watching" with 46 and "On Hold" with 38 titles, while no shows have been dropped.

Figure 27: Heatmap of the correlation matrix between days watches, mean scores, total entries, rewatched, episodes watched.

The correlation matrix heatmap shows various viewing metrics with key relationships highlighted, such as a strong correlation between "Episodes Watched" and "Days Watched" at

0.278. The matrix also indicates perfect correlations of 1.0 within metrics like "Rewatched" and "Mean Score," denoting self-referential correlations.

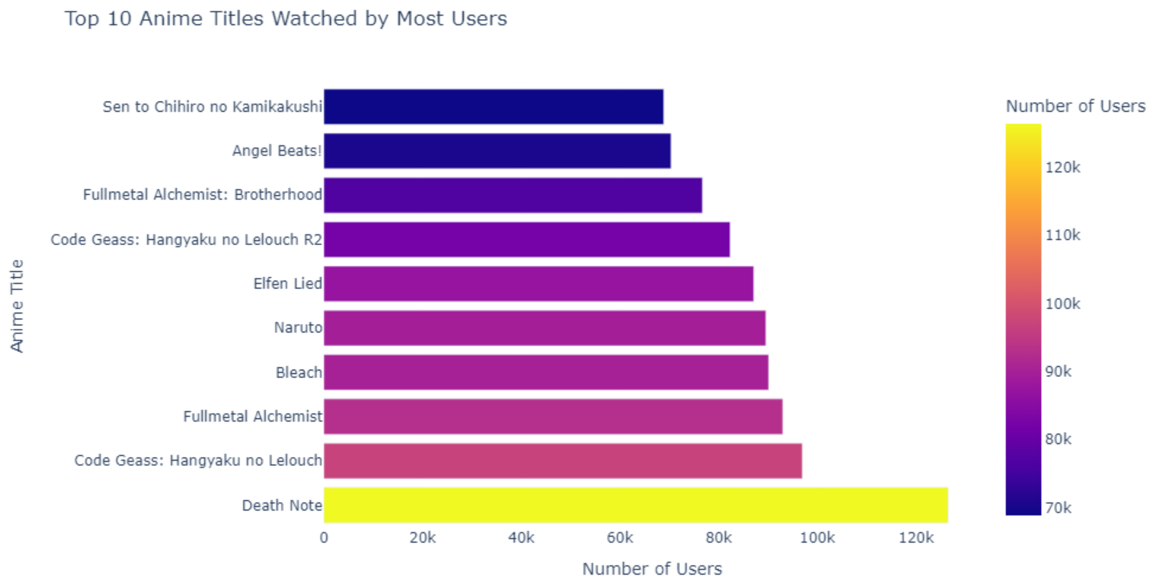

Figure 28: Bar plot of the anime titles watched by most users.

The bar chart shows the top 10 most-watched anime titles among users. "Death Note" leads with around 115,000 viewers, while "Sen to Chihiro no Kamikakushi" rounds out the list with just over 78,000 viewers. The chart uses a color gradient from yellow to purple to indicate the number of viewers, with purple representing higher viewership.

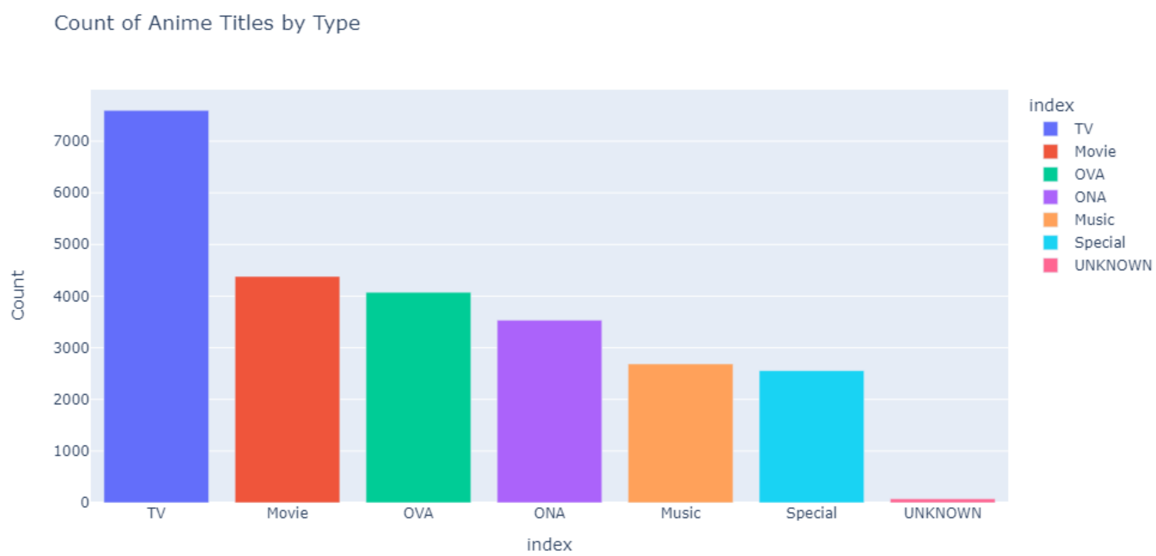

Figure 29: Bar plot of the anime title by types

The bar graph shows the distribution of anime titles by type, with "TV" having the highest count at over 7000 titles. Movies and OVAs also have substantial counts, around 4000 each, followed by ONAs and music-related titles. "Special" episodes and titles categorized as "UNKNOWN" are less frequent, indicating fewer entries in these categories.

2.4 Analysis Techniques

There are many different kinds of anime recommendation systems, and each one is designed to fit the interests and requirements of anime fans in a different way. The following is a list of different types of anime recommendation systems which have been used in the project work.

2.4.1 Collaborative Filtering

Collaborative filtering is a widely used technique in recommendation systems, including those for anime. It operates on the principle of leveraging the collective preferences and behaviors of users to make recommendations. In collaborative filtering for anime recommendation systems, there are typically two main approaches: user-based collaborative filtering and item-based collaborative filtering (What Is a Recommendation System?, n.d.).

In this work, a neural network-based model has been created for collaborative filtering, utilizing embeddings to represent users and animes in a lower-dimensional space, capturing their underlying preferences. The model has been trained using TensorFlow to predict user ratings for animes, with the optimization process ensuring that the model learns to recognize patterns and make accurate recommendations. With the trained model, similar animes and users can now be identified for generating recommendations.

The model summary is given below:

```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 user_encoded (InputLayer)      [(None, 1)]          0           []

 anime_encoded (InputLayer)     [(None, 1)]          0           []

 user_embedding (Embedding)     (None, 1, 128)       34564224    ['user_encoded[0][0]']

 anime_embedding (Embedding)    (None, 1, 128)       2112000     ['anime_encoded[0][0]']

 dot_product (Dot)              (None, 1, 1)         0           ['user_embedding[0][0]',
                                                                  'anime_embedding[0][0]']

 flatten (Flatten)              (None, 1)            0           ['dot_product[0][0]']

 dense (Dense)                  (None, 64)           128         ['flatten[0][0]']

 dense_1 (Dense)                (None, 1)            65          ['dense[0][0]']

==================================================================================================
Total params: 36,676,417
Trainable params: 36,676,417
Non-trainable params: 0
__________________________________________________________________________________________________
```

Figure 30: Deep Collaborative Filtering Model for Personalized Anime Recommendations.

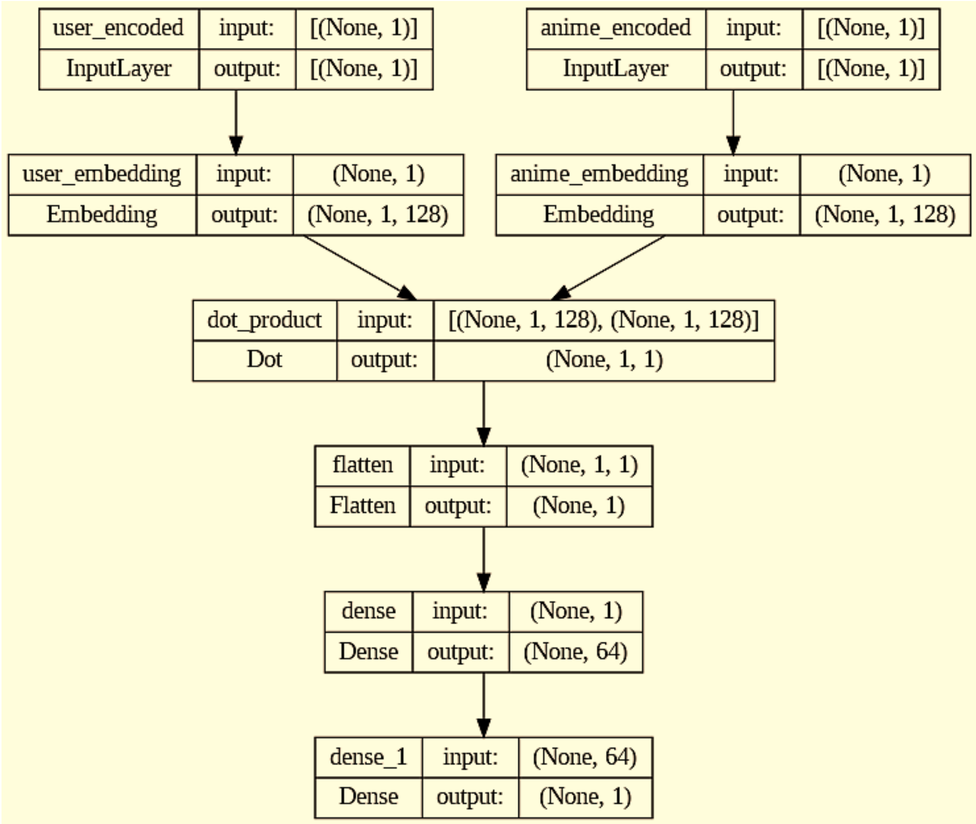

Figure 31: Anime Recommendation Model Architecture.

The model architecture clearly outlines the flow from input layers through embedding layers, interaction via a dot product, and subsequent processing through dense layers to produce the final output. This visual makes the operational structure of the model more understandable.

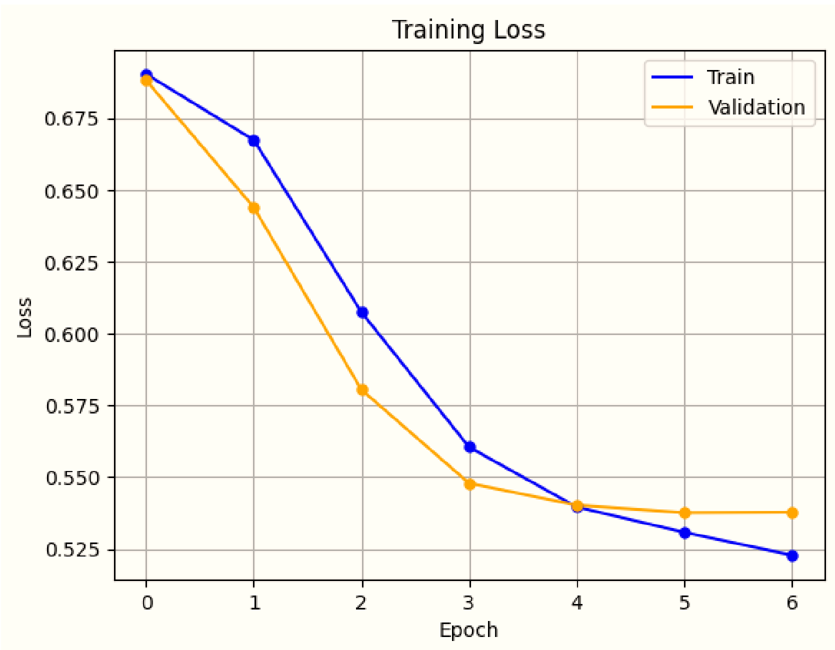

Figure 32: Loss vs Epoch graph for the deep learning model.

The training loss graph shows how both training and validation losses decrease significantly until about the fourth epoch, indicating that the model is initially learning and generalizing well. However, post the fourth epoch, while training loss continues to decline, validation loss begins to plateau and slightly increase, signaling the start of overfitting. An early stopping mechanism with a patience of three epochs is employed to prevent the model from overlearning the training data and losing generalization on new data. This mechanism ensures that training halts if the validation loss does not improve over three consecutive epochs, which is crucial for maintaining the model's effectiveness on unseen data. The graph underscores the importance of early stopping in model training to avoid overfitting and optimize model performance.

The model has been adjusted to recommend only those animes that have been rated by at least a certain number of users, which will be set as the threshold. This threshold is intended to ensure that the recommended anime titles have received a sufficient number of ratings, indicating a certain level of popularity or user engagement.

I. User-Based Collaborative Filtering:

- Similar User Identification: Similar users are identified based on their anime preferences. The find_similar_users function calculates user similarity using a weighted matrix and returns a data frame of similar users. It takes an input user, the number of similar users to find (n), and other optional parameters. The output is a list of similar users to the input user.
- Understanding User Preferences: This segment focuses on understanding the preferences of the selected user. The get_user_preferences function accepts a user ID as input and retrieves the anime preferences of that user. It considers the top-rated anime by the user and analyzes the genres they prefer. Additionally, the function provides an option to visualize the preferred genres using a word cloud. The output is a data frame containing the anime titles and their associated genres.
- Anime Recommendation for Users: In the final part, anime recommendations are made to the selected user based on the preferences of similar users. The get_recommended_animes function iterates through the list of similar users, retrieves their preferences, and identifies animes that are not present in the preferences of the selected user. Subsequently, it generates a list of recommended animes along with their genres and a brief synopsis. The output is a data frame containing the recommended animes for the user.

II. Item-Based Collaborative Filtering:

- In item-based collaborative filtering, recommendations are made based on the similarity between anime items themselves. The system identifies anime that are similar to those that a user has already liked or rated highly.
- The dot product of the weights matrix and the encoded index is computed to calculate the similarity between the given anime and all other anime.
- Once similar anime items are identified, the system recommends them to the user based on their similarity to items the user has already shown interest in.

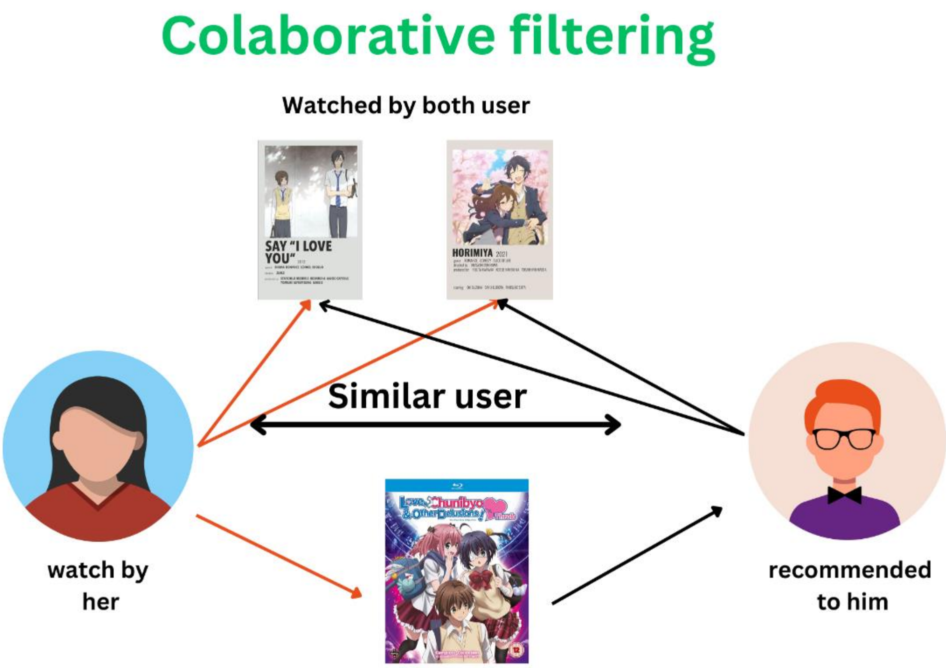

Figure 33: Collaborative Filtering.

Animes closest to Shingeki no Kyojin

	Name	Similarity	Genres	Synopsis
4	Death Note	84.29%	Supernatural, Suspense	Brutal murders, petty thefts, and senseless violence pollute the human world. In contrast, the realm of death gods is a humdrum, unchanging gambling den. The ingenious 17-year-old Japanese student Light Yagami and sadistic god of death Ryuk share one belief: their worlds are rotten.\n\nFor his own amusement, Ryuk drops his Death Note into the human world. Light stumbles upon it, deeming the first of its rules ridiculous: the human whose name is written in this note shall die. However, the temptation is too great, and Light experiments by writing a felon's name, which disturbingly enacts his first murder.\n\nAware of the terrifying godlike power that has fallen into his hands, Light—under the alias Kira—follows his wicked sense of justice with the ultimate goal of cleansing the world of all evil-doers. The meticulous mastermind detective L is already on his trail, but as Light's brilliance rivals L's, the grand chase for Kira turns into an intense battle of wits that can only end when one of them is dead.
3	Attack on Titan Season 2	83.19%	Action, Drama, Suspense	For centuries, humanity has been hunted by giant, mysterious predators known as the Titans. Three mighty walls—Wall Maria, Rose, and Sheena—provided peace and protection for humanity for over a hundred years. That peace, however, was shattered when the Colossal Titan and Armored Titan appeared and destroyed the outermost wall, Wall Maria. Forced to retreat behind Wall Rose, humanity waited with bated breath for the Titans to reappear and destroy their safe haven once more.\n\nIn Shingeki no Kyojin Season 2, Eren Yeager and others of the 104th Training Corps have just begun to become full members of the Survey Corps. As they ready themselves to face the Titans once again, their preparations are interrupted by the invasion of Wall Rose—but all is not as it seems as more mysteries are unraveled. As the Survey Corps races to save the wall, they uncover more about the invading Titans and the dark secrets of their own members.
2	Code Geass: Lelouch of the Rebellion R2	81.95%	Action, Award Winning, Drama, Sci-Fi	One year has passed since the Black Rebellion, a failed uprising against the Holy Britannian Empire led by the masked vigilante Zero, who is now missing. At a loss without their revolutionary leader, Area 11's resistance group—the Black Knights—find themselves too powerless to combat the brutality inflicted upon the Elevens by Britannia, which has increased significantly in order to crush any hope of a future revolt. \n\nLelouch Lamperouge, having lost all memory of his double life, is living peacefully alongside his friends as a high school student at Ashford Academy. His former partner C.C., unable to accept this turn of events, takes it upon herself to remind him of his past purpose, hoping that the mastermind Zero will rise once again to finish what he started, in this thrilling conclusion to the series.
1	Attack on Titan Season 3 Part 2	81.85%	Action, Drama	Seeking to restore humanity's diminishing hope, the Survey Corps embark on a mission to retake Wall Maria, where the battle against the merciless "Titans" takes the stage once again.\n\nReturning to the tattered Shiganshina District that was once his home, Eren Yeager and the Corps find the town oddly unoccupied by Titans. Even after the outer gate is plugged, they strangely encounter no opposition. The mission progresses smoothly until Armin Arlert, highly suspicious of the enemy's absence, discovers distressing signs of a potential scheme against them. \n\nShingeki no Kyojin Season 3 Part 2 follows Eren as he vows to take back everything that was once his. Alongside him, the Survey Corps strive—through countless sacrifices—to carve a path towards victory and uncover the secrets locked away in the Yeager family's basement.
0	ERASED	80.92%	Mystery, Supernatural, Suspense	When tragedy is about to strike, Satoru Fujinuma finds himself sent back several minutes before the accident occurs. The detached, 29-year-old manga artist has taken advantage of this powerful yet mysterious phenomenon, which he calls "Revival," to save many lives.\n \nHowever, when he is wrongfully accused of murdering someone close to him, Satoru is sent back to the past once again, but this time to 1988, 18 years in the past. Soon, he realizes that the murder may be connected to the abduction and killing of one of his classmates, the solitary and mysterious Kayo Hinazuki, that took place when he was a child. This is his chance to make things right.\n \nBoku dake ga Inai Machi follows Satoru in his mission to uncover what truly transpired 18 years ago and prevent the death of his classmate while protecting those he cares about in the present.

(a)

> Top recommendations for user: 1213175

	n	anime_name	Genres	Synopsis
0	22486	Ookami to Koushinryou II	Adventure, Fantasy, Romance	Traveling merchant Kraft Lawrence continues his northward journey with wolf goddess Holo, in search of her lost home of Yoitsu. Lawrence and his sharp-witted partner continue to make some small profits along the way, while slowly uncovering more information about Holo's hometown. However, the road to Yoitsu is a bumpy one filled with many troubles—Lawrence runs into a charming young fellow merchant who has his eyes set on the female wolf companion, and he begins to doubt if Holo will remain by his side; he and the goddess will also have to consider precarious and risky business deals as Lawrence strives to achieve his dream of becoming a shopowner. All the while, with his determination tested at every turn during his journey, Lawrence must question his relationship with Holo, take on business ventures, and ask himself whether it is time for him and Holo to go their separate ways.
1	20864	Ga-Rei: Zero	Action, Supernatural, Suspense	In Japan, there exists a government agency known as the Supernatural Disaster Countermeasures Division (SDCD), whose duty is to protect the citizens from creatures unseen. They are able to dispatch these monsters swiftly and without alerting the general public. But currently, they face a different challenge: the betrayal of one of their own.\n\nAfter the death of her mother several years ago, Kagura Tsuchimiya has been fostered by the Isayama family and forms a close sister-like bond with their daughter Yomi. The two become inseparable, and together they work for the SDCD as highly skilled exorcists. However, as the stress and consequences of their sacred duty weigh on them both, and family politics come into play, Kagura and Yomi begin to slowly drift apart. One of them grows earnestly into her role as an exorcist, and the other heads down a dark path from which there may be no redemption.
2	36094	Mirai Nikki (TV)	Action, Supernatural, Suspense	Yukiteru Amano is a shy middle schooler who regularly keeps track of what he does in his daily life by writing down all of his activities on his phone—a digital diary. Despite having no friends at school, Yukiteru is frequently seen talking to his supposedly imaginary friends Deus Ex Machina, the god of time and space; and Deus' servant, Mur Mur.\n\nOne day, Yukiteru wakes up and discovers that certain events of his day are preemptively displayed on his cellphone. While initially dismissing it as a coincidence, he slowly realizes that the incidents written in his phone actually take place in the near future. After spending the day benefiting from this new asset, Yukiteru learns that his classmate Yuno Gasai possesses a similar diary.\n\nAs the two team up to defeat an odd pursuer and head back to their respective homes, Deus Ex Machina explains that they—alongside 10 other contestants—have been drawn into a survival game whose victor will become the deity's successor. With no other options, Yukiteru and Yuno must use their cellphones—now called "Future Diaries"—to survive this unforgiving battle royale.
3	23222	Danshi Koukousei no Nichijou	Comedy	Roaming the halls of the all-boys Sanada North High School are three close comrades: the eccentric ringleader with a hyperactive imagination Hidenori, the passionate Yoshitake, and the rational and prudent Tadakuni. Their lives are filled with giant robots, true love, and intense drama... in their colorful imaginations, at least. In reality, they are just an everyday trio of ordinary guys trying to pass the time, but who said everyday life couldn't be interesting? Whether it's an intricate RPG reenactment or an unexpected romantic encounter on the riverbank at sunset, Danshi Koukousei no Nichijou is rife with bizarre yet hilariously relatable situations that are anything but mundane.
4	10129	Berserk: Ougon Jidai-hen I - Haou no Tamago	Action, Adventure, Drama, Fantasy, Horror	In the Kingdom of Midland, a mercenary named Guts wanders the land, preferring a life of conflict over a life of peace. Despite the odds never being in his favor, he is an unstoppable force that overcomes every opponent, wielding a massive sword larger than himself.\n\nOne day, Griffith, the mysterious leader of the mercenary group Band of the Hawk, witnesses the warrior's battle prowess and invites the wandering swordsman to join his squadron. Rejecting the offer, Guts challenges Griffith to a duel—and, much to the former's surprise, is subsequently defeated and forced to join.\n\nNow, Guts must fight alongside Griffith and his crew to help Midland defeat the Empire of Chuder. However, Griffith seems to harbor ulterior motives, desiring something much larger than just settling the war...
5	7531	Berserk: Ougon Jidai-hen III - Kourin	Action, Adventure, Drama, Fantasy, Horror	The Band of the Hawk has dwindled in the year since Guts left them on his journey to forge his own destiny. Unaware of their fate, Guts returns to the Hawks—now being led by his former ally Casca—after a rumor about them passes his way. Once the saviors of the kingdom of Midland, the Band of the Hawk are now hunted as they desperately fight for their lives while plotting to free their leader, Griffith, after he was imprisoned for committing treason. But the man they save is far from the Griffith they remember.\n\nGriffith is a shell of his former charismatic self after a year of continuous, horrific torture. No longer able to walk, speak, or even hold a sword, he has nothing but the small, strange trinket, the Crimson Behelit, that will not leave him. The entire Band of the Hawk want to rise to greatness once more, but how much are they willing to sacrifice to return to their past glory? It doesn't seem possible, but when Griffith's heart darkens and a solar eclipse blackens the sky, the Behelit offers a choice that will leave the Band of the Hawk with a blood-soaked fate that will haunt them for the rest of their days.

(b)

Figure 34: Output of Collaborative Filtering (a) Item Based: Top 5 Recommendation for "Shingeki no Kyojin" (b) Used Based: Top 5 Recommendation for random user 1213175.

2.4.2 Content Based Filtering

Content-based filtering is another popular technique used in recommendation systems, including those for anime. It operates on the principle of recommending items that are similar to those that a user has liked or interacted with in the past. In the context of anime recommendation systems, content-based filtering focuses on analyzing the content features of anime (such as genres, synopsis, and themes) to make recommendations (What Is a Recommendation System?, n.d.).

- Feature Extraction: TF-IDF vectorization has been utilized to create a TF-IDF matrix for anime genres, quantifying the importance of genres in each anime's description. This process has involved encoding the genre descriptions into numerical vectors based on their frequency and inverse document frequency.

$$TF - IDF(t, d) = TF(t, d) * \log\left(\frac{N}{DF(t)}\right) \tag{2}$$

$where,$

  - $TF(t, d)$ as the term frequency of term t in document d (anime genre description).
  - $DF(t)$ as the document frequency of term t (number of anime descriptions containing term t).
  - N as the total number of anime descriptions.

- Similarity Calculation: Cosine similarity is then computed between animes based on their genre descriptions, enabling the determination of their similarity. This involves comparing the TF-IDF vectors of different anime to measure the cosine of the angle between them, indicating their degree of similarity in terms of genre content.

$$Cosine\ Similarity\ (a, b) = \frac{a\,.b}{\|a\|\,\|b\|} = \frac{a\,.b}{\sqrt{\sum_{i=1}^{n} a_i^2}\,\sqrt{\sum_{i=1}^{n} b_i^2}} \tag{3}$$

$where,$

  - $a\,.\,b$ represents the dot product (also known ad inner product) between vectors a and b.
  - $\|a\|$ and $\|b\|$ denotes the Euclidean norm (magnitude) of vector a and b respectively.

This formula yields a value between -1 and 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect dissimilarity. In the context of anime recommendation systems, cosine similarity is commonly used to measure the similarity between the genre vectors of different anime.

- Recommendation Generation: Finally, leveraging the computed similarity scores and ratings, content-based recommendations are generated. Animes that are deemed similar to a given anime are recommended, taking into account their genre and score. This recommendation process utilizes the cosine similarity scores to identify animes with similar genre descriptions, supplemented by user ratings to prioritize recommendations based on user preferences and satisfaction.

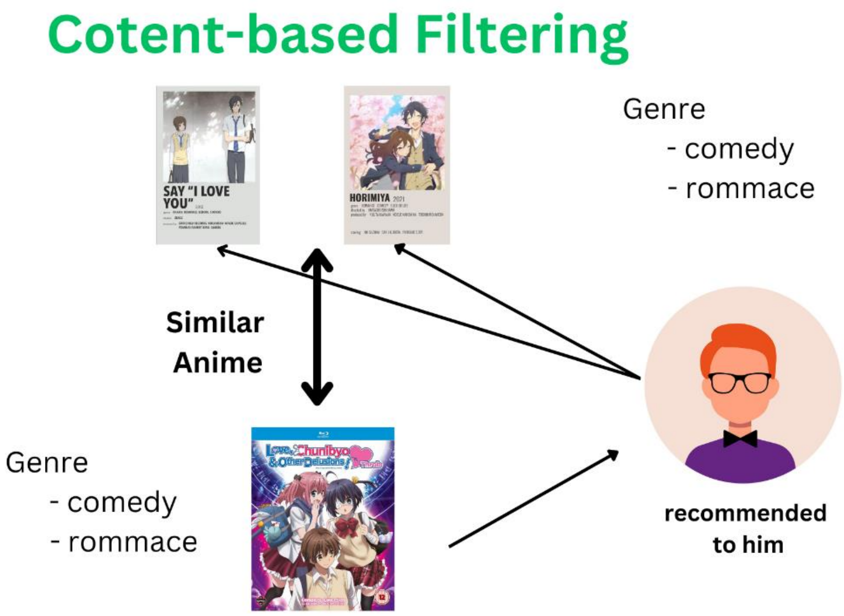

Figure 35: Content Based Filtering.

Recommendations for "Shingeki no Kyojin":

	Name	Genres	Score
5873	Mahou Shoujo Madoka★Magica	Award Winning, Drama, Suspense	8.36
3271	Evangelion: 3.0+1.0 Thrice Upon a Time	Action, Award Winning, Drama, Sci-Fi, Suspense	8.61
980	Sentou Yousei Yukikaze	Action, Award Winning, Drama, Sci-Fi, Suspense	7.2
6645	Mahou Shoujo Madoka★Magica Movie 3: Hangyaku no Monogatari	Award Winning, Drama, Mystery, Suspense	8.47
21814	Ikuta no Kita	Adventure, Award Winning, Fantasy, Suspense	6.22
2469	Hidamari no Ki	Action, Award Winning, Drama	7.06
4135	Birthday Boy	Action, Award Winning, Drama	5.65
24	Koukaku Kidoutai	Action, Award Winning, Mystery, Sci-Fi, Suspense	8.27
20	Neon Genesis Evangelion	Action, Avant Garde, Award Winning, Drama, Sci-Fi, Suspense	8.35
9854	Koe no Katachi	Award Winning, Drama	8.94

Figure 36: Output for Content based filtering: Top 10 Content Based Recommendation for "Shingeki no Kyojin".

2.4 Technologies and Tools

This project extensively utilizes Python programming language version 3.10.12 in conjunction with the Pandas library tool version 2.0.3 and NumPy version 1.23.5. These widely recognized software libraries are renowned for their effectiveness in data analysis tasks, making it a pivotal component of our research endeavor. To gain valuable insights (EDA), Plotly has been utilized, a fantastic library that provides us with interactive and engaging visualizations. The operational tasks are exclusively conducted within the Google Colab Pro environment, which boasts a robust hardware configuration. Specifically, the Colab Pro setup encompasses a TPU with 8 cores, offering a disk size of 225.8GB and high RAM capacity of 12.7GB. Moreover, all deep learning operations are executed using the TensorFlow framework version 2.15.0, ensuring compatibility and optimal performance across the board.

3. Case Study / Practical Application

A real-world application of a recommendation system in the anime domain, let's look at Netflix's use of big data for its anime recommendation engine.

3.1 Relevant Case Study: Netflix's Recommendation System

Background: Netflix, a global streaming service, offers a vast array of entertainment content, including a significant collection of anime. It uses advanced recommendation systems to personalize content offerings, improving viewer satisfaction and engagement.

3.2 Application of Big Data Analysis

Data Collection: Netflix collects extensive data about user interactions with their service, including viewing history, search queries, ratings provided by users, and the time spent on each show. For anime specifically, they also gather data on subtitles selected and audio tracks used, which can indicate preferences for localized content or original Japanese audio.

Analysis Techniques: Netflix employs a variety of machine learning algorithms and techniques for its recommendation systems. A significant approach is the use of deep learning to analyze not just user behavior but also content features. They utilize Convolutional Neural Networks (CNNs) to analyze visual elements from anime, such as animation style and color themes, and Natural Language Processing (NLP) techniques to understand and categorize plot summaries and viewer reviews.

Technology Stack: The system likely utilizes big data tools such as Apache Kafka for handling real-time data streams, Apache Cassandra for scalable data storage, and potentially machine learning frameworks like TensorFlow for model development. The architecture is supported by a robust cloud infrastructure to handle the vast scale of data processing and analysis required.

3.3 Impact and Outcomes

Enhanced Personalization: Through its sophisticated use of big data, Netflix can offer highly personalized anime recommendations, catering to the nuanced preferences of a diverse global audience. This includes suggesting anime based on indirect preferences inferred from non-anime viewing habits.

Global Reach: By effectively analyzing subtitle and audio track preferences, Netflix can better understand regional tastes, helping inform decisions about which anime series to license and offer in different markets.

Viewer Retention: The personalized and engaging experience provided by effective recommendation systems helps Netflix maintain a high level of viewer retention and reduce subscription cancellations.

Content Strategy: Insights gained from data analysis influence Netflix's content strategy, not just in licensing, but also in original content production. Understanding popular genres and themes within the anime community can guide Netflix in producing original anime that appeals to both hardcore fans and casual viewers.

This case study highlights the strategic use of big data in enhancing content discoverability and viewer satisfaction, playing a crucial role in Netflix's success in the competitive streaming market.

4. Conclusion

The anime recommendation system developed incorporates collaborative filtering (both user-based and item-based) and content-based filtering techniques to provide personalized recommendations to users. Through the implementation of these techniques, several key findings and outcomes have emerged. Firstly, collaborative filtering has significantly improved user engagement by analyzing user ratings and viewing history to suggest anime that align with individual preferences and interests. This has led to increased user satisfaction and longer durations of platform usage. Additionally, the utilization of both user-based and item-based collaborative filtering has resulted in a more diverse range of anime recommendations, exposing users to a broader spectrum of content and enhancing their viewing experience. Content-based filtering has further enhanced the system's ability to predict user preferences accurately by recommending anime based on characteristics such as genre, themes, and animation style. This approach ensures that recommendations are tailored to specific user tastes, leading to higher relevance and acceptance of suggested anime. Furthermore, the incorporation of content-based filtering has mitigated the cold-start problem, allowing the system to make relevant suggestions even for users with limited interaction history.

In our future endeavors, we aim to enhance our recommendation system by implementing a hybrid approach that combines collaborative filtering and content-based filtering techniques, integrating user feedback and reviews for improved accuracy. Additionally, we plan to explore advanced deep learning models like RNNs and transformers to further enhance recommendation performance. Our system will evolve to provide real-time updates based on user interactions and will leverage external data sources such as user demographics and anime-related news for personalization. Sentiment analysis on anime reviews will offer insights into audience preferences, while user clustering will enable more targeted recommendations and marketing strategies. A user-friendly web interface will facilitate exploration of recommendations and detailed anime information, with social media integration for sharing favorites. Time series analysis will uncover trends in anime popularity, while personalized watchlists and sentiment-based filtering will cater to individual preferences and emotions.

5. References

Anime Dataset 2023. (2023, July 28). Kaggle. https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset

Ekstrand, M. D. (2011). Collaborative Filtering Recommender Systems. *Foundations and Trends® in Human–Computer Interaction*, *4*(2), 81–173. https://doi.org/10.1561/1100000009

Ko, H., Lee, S., Park, Y., & Choi, A. (2022, January 3). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*, *11*(1), 141. https://doi.org/10.3390/electronics11010141

Kumar, P., & Thakur, R. S. (2018, April 7). Recommendation system techniques and related issues: a survey. *International Journal of Information Technology*, *10*(4), 495–501. https://doi.org/10.1007/s41870-018-0138-8

Singhal, A., Sinha, P., & Pant, R. (2017, December 15). Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works. International Journal of Computer Applications, 180(7), 17–22. https://doi.org/10.5120/ijca2017916055

What is a Recommendation System? (n.d.). NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/recommendation-system/#:~:text=A%20recommendation%20system%20

Work Distribution

Name	ID	Email	Contribution	Percentage	Signature
Kowshik Sankar Roy	20231040	kowshikroy777@gmail.com	Conceptualization, Software, Writing: Methodology	25%	
Koushik Paul	20231023	paulkoushik9400@gmail.com	Writing: Introduction, Case Study	20%	
Bashirul Alam	20231030	amitalam1243@gmail.com	Writing: EDA, Original Draft Preparation, Referencing	25%	
Sumaiya Binte Haque	20231081	sumaiyasathi11@gmail.com	Presentation Slide Preparation	20%	
Md. Ariful Islam	20231043	ariful.islameee@gmail.com	Writing: Conclusion	10%	