FINAL PROJECT REPORT

PROJECT TITLE: ANIME RECOMMENDATION

FOUNDATION OF DATA ANALYTICS (CSE3505)

FALL SEMESTER 2022-23

SUBMITTED BY:

VENKAT REDDY	19MIA1104
MOUNIKA SAI	19MIS1013
GAYATHRI	19MIS1035
SIRI MAHALAKSHMI	19MIS1044
MANASA	19MIS1063
MOHAN	19MIS1200

SUBMITTED TO:

VERGIN RAJA SAROBIN M

ABSTRACT

- Our project aims to Build a better anime recommendation system which recommends you the top rated anime in the data set based on user viewing and rating history.

- We Built a better anime recommendation system based only on user viewing history.

- With the data, we might cluster and predict the data set using data mining algorithms.

- We are going to infer the best and top rated anime and recommend them to the user by using this data set.

- Each anime has its own characteristic which complies with specific user’s interest. Therefore, a personalization engine was needed to provide a recommendations.

1.INTRODUCTION

Here we have collected the anime dataset from Kaggle , and did prediction methods for ratings and genre based on the type of broadcast and cluster analysis with partitioning methods.

- we did data dispersion for statistical view and for visual analysis we used histograms and boxplots for ratings. system listed most popular Anime genres and based on that we infer top anime recommendations for gerne and types(movies, ova, tv).
- Using genre and type it finds the similarity between animes and recommend it to the user.
- system will recommend more animes for the user, based on user rating and the genres they watch frequently. it filters and recommend animes which are top rated on different genres and types(movies, ova, tv).
- Our project Give recommendations based on the preferences of other similar users
- Give recommendations based on the similarity between items and the rating history of the user.

PROBLEM STATEMENT:

Each piece of streaming content has its own audience and ratings. If viewers enjoy the content, they give it positive reviews. Where, though, does it apply? Viewers can lose hours scrolling through hundreds, and occasionally even thousands, of anime episodes without ever finding anything they enjoy. To improve streaming and increase revenue and time spent on a website, businesses must be given recommendations based on their preferences and needs.

2.LITERATURE SURVEY

1.RECOMMENDATION SYSTEM FOR ANIME USING MACHINE LEARNING ALGORITHMS:

The existing model's potential is restricted due to the lack of content-based filtering technique. The system is implemented in multiple ways. They intended to address Content Based Filtering, Popularity Based Recommendation System, Collaborative Filtering using KNN and Collaborative Filtering using SVD approaches in their research work.

2.ANIME RECOMMENDATION SYSTEM:

Utilizing datasets and algorithms, recommender systems create new possibilities for finding personalized information. It also enables users to access goods and services that are not immediately available to users on the system, which helps to relieve the issue of information overload that is a very regular occurrence with information retrieval systems. The two were covered in this project.

3.A DEEP LEARNING RECOMMENDER SYSTEM FOR ANIME:

This Recommender system can help to generate personalized and accurate anime recommendations for users on platforms. This research paper will help to build a recommender system with a Deep Learning collaborative filtering-based model which provides anime recommendations. This research paper Attempts to solve the problems in conventional Recommender Systems such as the data-sparsity and cold-start.

4.RIKONET: A NOVEL ANIME RECOMMENDATION ENGINE:

Today, anime is very popular, especially among the younger generations. There are numerous show genres available, which is drawing in a rising number of viewers. In this effort, they have created a novel hybrid recommendation system that might serve as both a way to discover new anime genres and titles as well as a mechanism for making recommendations.

5.AN INTELLIGENT ANALYTICAL FRAMEWORK FOR ANIME RECOMMENDATION AND PERSONALIZATION USING AI AND BIG DATA ANALYSIS:

Two experiments were run. They initially tested if the training set's popularity had an impact on overall accuracy. Because of the nature of SVM and the popularity of the top 100 anime, they discovered that employing them resulted in poorer accuracy. In their second experiment they carried trials to determine whether common and uncommon genres had an effect on accuracy and the results showed that having uncommon genres will yield better accuracy.

6.COLLABORATIVE RECOMMENDATION SYSTEM IN USERS OF ANIME FILMS:

The anime movie is one of the most well-liked items on the list. In this instance, they carry out research to suggest anime movies based on reviews of previously viewed movies. A method called collaborative filtering involves tallying up forecasts, recommendations, and similarities.

The collaborative filtering method judges the quality of the user's data based on how closely their preferences match those of other users. Alternating Least Squares (ALS) is used to build the recommendation system using the training data.

3.PROPOSED WORK

3.1 DATA SOURCE

The film dataset used in our experiment was obtained from the anime recommendation database “Kaggle”. Data provide two files namely anime data and rating on anime. This data consists of a relation between 73,516 users and 12,294 anime. Each user is able to add anime to their completed list and give it a rating and this data set is a compilation of those ratings.

Anime.csv

anime_id - myanimelist.net's unique id identifying an anime.

name - full name of anime.

genre - comma separated list of genres for this anime.

type - movie, TV, OVA, etc.

episodes - how many episodes in this show. (1 if movie).

rating - average rating out of 10 for this anime.

members - number of community members that are in this anime's "group".

Rating.csv

user_id - non identifiable randomly generated user id.

anime_id - the anime that this user has rated.

rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

3.2 DATA ANALYTICS MODELS

➢ COLLOBRATIVE FILTERING

This system provides recommendations based on the tastes of other users who share similar likes. Similarities between people and objects are simultaneously used in collaborative filtering to provide suggestions. This enables suggestions, allowing the collaborative filtering models to suggest an item to user A based on the interests of a user B who shares those interests. We'll develop a clear recommendation system based on clear historical information, in this example, "ratings" that users have provided for each movie. Rating df will be transformed into a 2-dim matrix with user-id in the columns and anime-id in the rows before being saved in a sparse manner. reducing the amount of time it takes to fit the model.

➢ CONTENT BASED FILTERING

In order for a content-based recommender to function, we must collect data from the user, either explicitly (via ratings) or implicitly (clicking on a link). By using the information, we can build a profile of the user, which is then used to make suggestions to the user. As the user adds more information or acts more frequently on the recommendation, the engine gets more accurate.

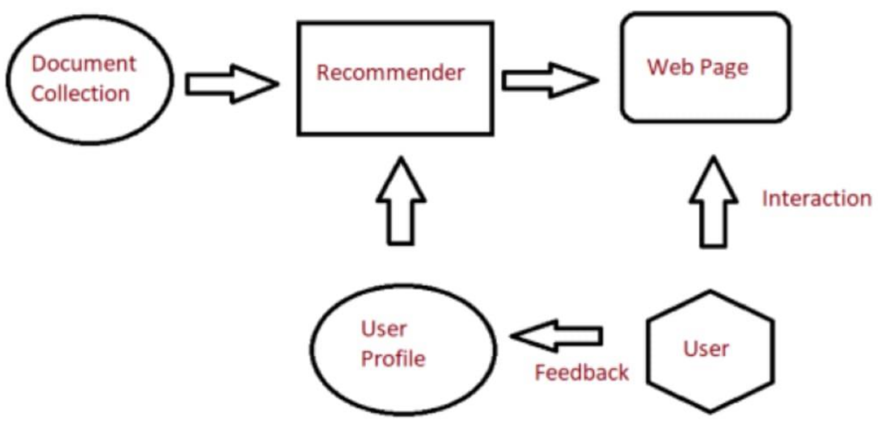
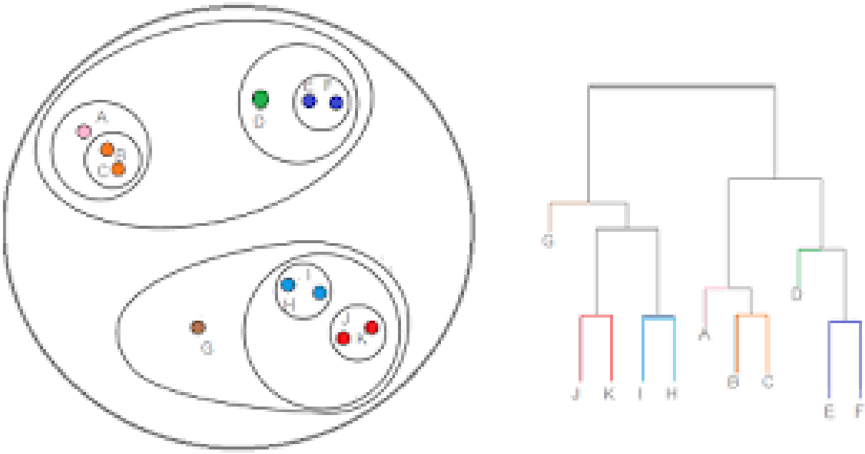


Fig: Recommender System

➢ HIERARCHICAL-AGGLOMERATIVE

One well-known HCA is the agglomerative hierarchical clustering algorithm. The bottom-up methodology is used to cluster the datasets. This means that this algorithm starts by treating each dataset as a single cluster and then begins combining the two clusters that are the closest to one another. It continues doing this until every cluster has been combined into a single cluster that has every dataset.



➢ COSINE SIMILARITY

The cosine similarity index calculates how similar two vectors in an inner product space are to one another. It establishes whether two vectors are roughly pointing in the same direction by calculating the cosine of the angle between them. The mathematical definition of cosine similarity is the vectors' dot product divided by their magnitude. For instance, the similarity between two vectors A and B is determined as

$$similarity(A, B) = cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

➢ DATA DISPERSION:

Spreading out the data set is called data dispersion. and that the measurement of data quantity is a major tendency. When a data collection contains a significant value, the values are dispersed widely throughout the set. From a data mining perspective, standpoint, we must consider how enormous data sets can be efficiently computed databases. Therefore, we determined central trends and depicted them in Rating, type, and member attribute histograms. Bivariate analysis can be useful in evaluating straightforward association hypotheses. It specifies how the ratings of the attributes relate to the categories and episodes. Standard deviation, variance, and cumulative are examples of mixed measurements percentage.

4. RESULTS

DATA DISPERSION:

[101]:

```
print(f"Summary of The Rating Dataset :")
rating.describe().T
```

Summary of The Rating Dataset :

	count	mean	std	min	25%	50%	75%	max
user_id	7813737.0	36727.956745	20997.946119	1.0	18974.0	36791.0	54757.0	73516.0
anime_id	7813737.0	8909.072104	8883.949636	1.0	1240.0	6213.0	14093.0	34519.0
rating	7813737.0	6.144030	3.727800	-1.0	6.0	7.0	9.0	10.0

Summary of rating dataset

[96]:

```
print(f"Summary of The Anime Dataset :")
anime.describe().T
```

Summary of The Anime Dataset :

	count	mean	std	min	25%	50%	75%	max
anime_id	12294.0	14058.221653	11455.294701	1.00	3484.25	10260.50	24794.50	34527.0
rating	12064.0	6.473902	1.026746	1.67	5.88	6.57	7.18	10.0
members	12294.0	18071.338864	54820.676925	5.00	225.00	1550.00	9437.00	1013917.0

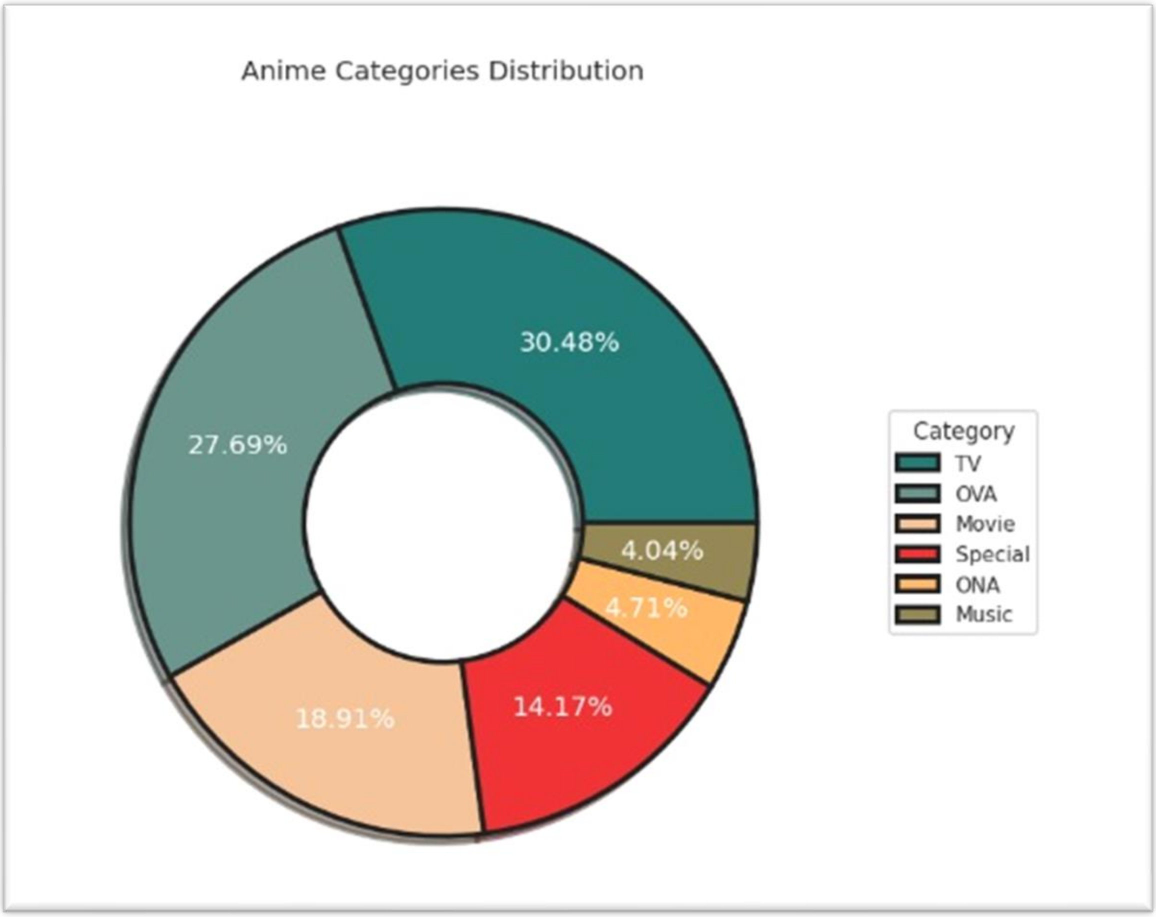
+ Code + Markdown

[97]:

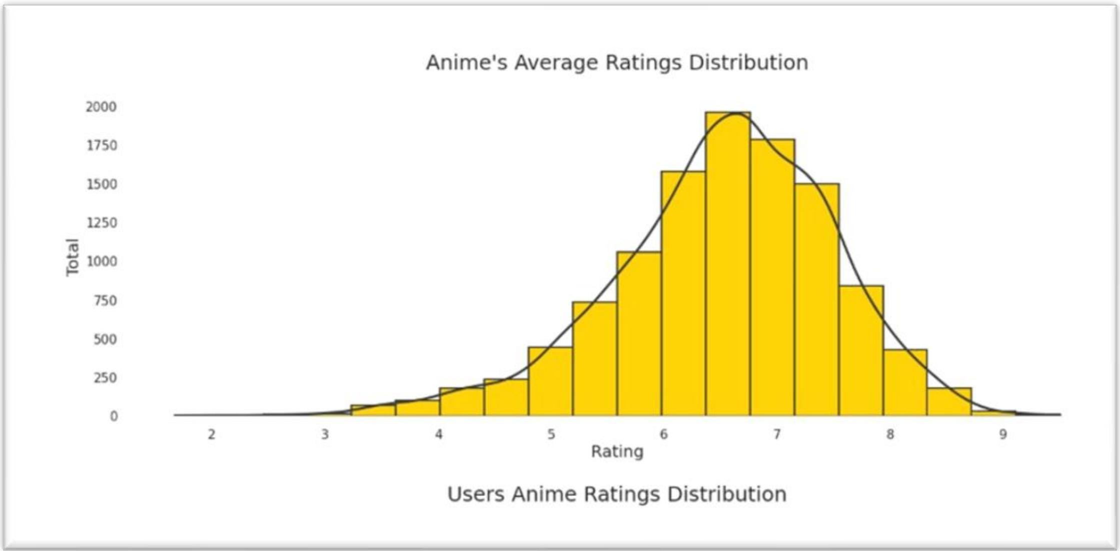
```
anime.describe(include=object)
```

	name	genre	type	episodes
count	12294	12232	12269	12294
unique	12292	3264	6	187
top	Shi Wan Ge Leng Xiaohua	Hentai	TV	1
freq	2	823	3787	5677

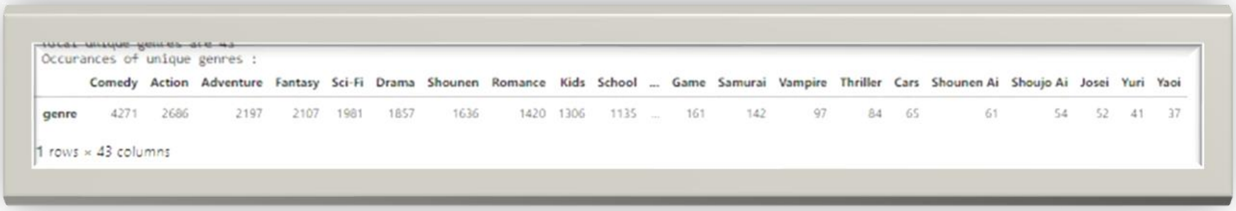
Summary of anime dataset



Platforms where anime is viewed by users



Overall view of anime rating given by users

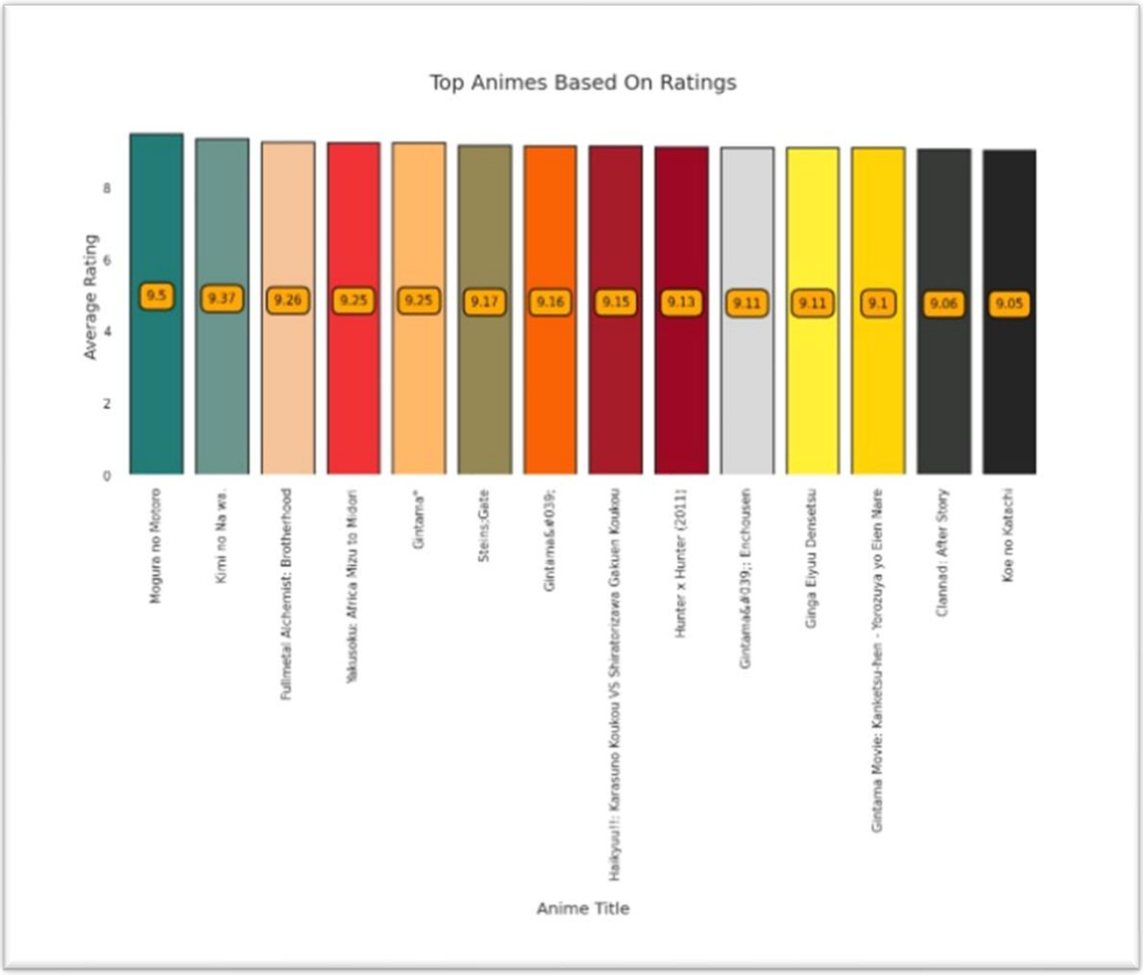


Occurances of unique genres :

	Comedy	Action	Adventure	Fantasy	Sci-Fi	Drama	Shounen	Romance	Kids	School	...	Game	Samurai	Vampire	Thriller	Cars	Shounen Ai	Shoujo Ai	Josei	Yuri	Yaoi
genre	4271	2686	2197	2107	1981	1857	1636	1420	1306	1135	...	161	142	97	84	65	61	54	52	41	37

1 rows × 43 columns

Genre of anime in dataset



Most viewed anime in dataset

COLLOBRATIVE FILTERING:

It Gives recommendations based on the preferences of other similar users.

```
Recommendations for Ahiru Rikusentai viewers :
```

Out[132]:

	Anime Name	Rating
No		
1	Tekusuke Monogatari	4.00
2	Nobara	4.59
3	Kanimanji Engi	4.79
4	Norakuro Ittouhei	4.44
5	Kinken Chochiku Shiobara Tasuke	4.21

CONTENT BASED FILTERING:

```
Recommendations for Naruto viewers :
```

Out[135]:

	Anime Name	Rating
No		
1	To LOVE-Ru Darkness OVA	7.82
2	Hanbun no Tsuki ga Noboru Sora	7.69
3	Mai-HiME	7.59
4	Doraemon Movie 28: Nobita to Midori no Kyojin Den	7.54
5	Rurouni Kenshin Special	7.51
6	Pikmin Short Movies	7.27
7	Deadman Wonderland OVA	7.12
8	Anata mo Robot ni Nareru feat. Kamome Jidou Ga...	5.12
9	Shinpi no Hou	5.37
10	Toaru Majutsu no Index: Endymion no Kiseki	7.71

Recommendations for naruto viewers

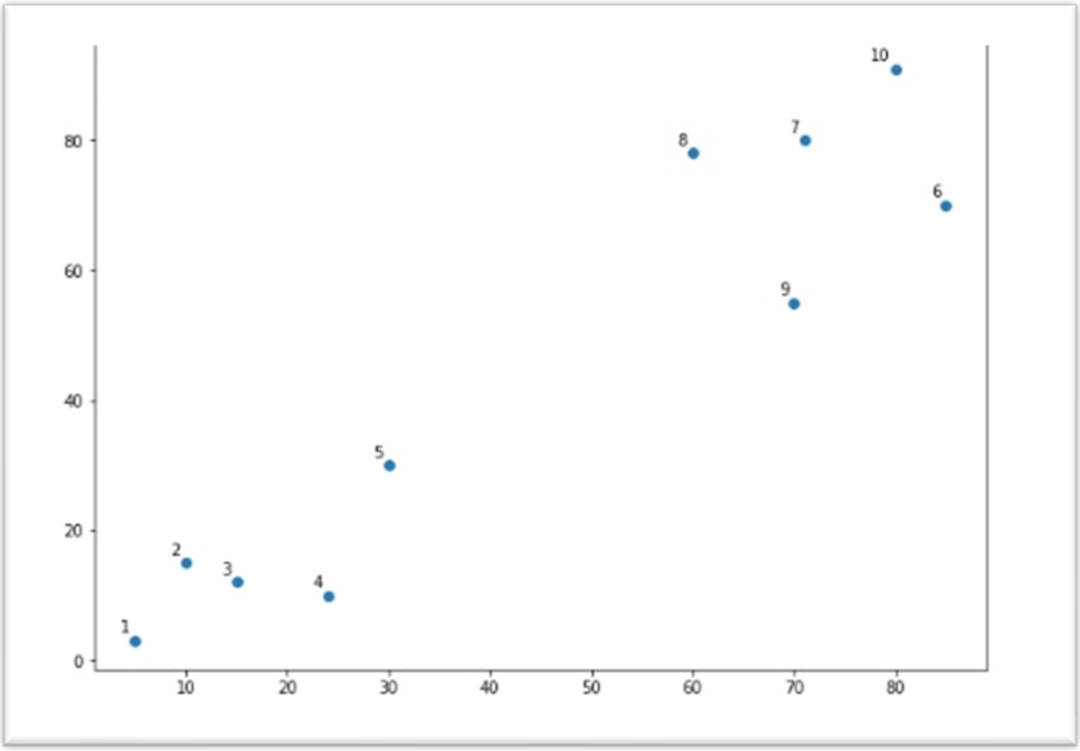
```
Recommendations for Death Note viewers :
```

Out[136]:

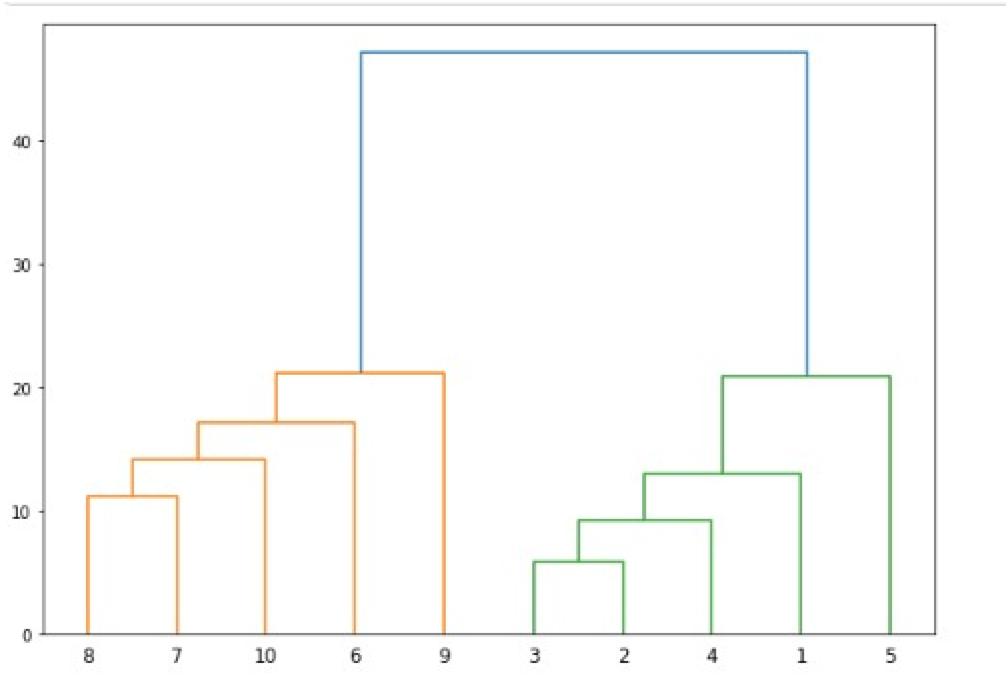
	Anime Name	Rating
No		
1	Hachimitsu to Clover Specials	7.85
2	Trapp Ikka Monogatari	7.75
3	Major S1	8.42
4	Hakkenden: Touhou Hakken Ibun	7.57
5	Ushi Atama	4.87
6	ef: A Tale of Melodies.	8.18
7	Saki Achiga-hen: Episode of Side-A Specials	7.63
8	One Piece: Oounabara ni Hirake! Dekkai Dekkai ...	7.43
9	Kizumonogatari II: Nekketsu-hen	8.73
10	Gundam Evolve	6.89

Recommendations for Death Note viewers

HIERARCHICAL-AGGLOMERATIVE:



we can see two clusters: the first at the bottom left consisting of points 1-5 while the second at the top right consisting of points 6-10.



Dendrogram

COSINE SIMILARITY:

```
Anime missing values (%):

rating      1.87
genre       0.50
type        0.20
anime_id    0.00
name        0.00
episodes    0.00
members     0.00
dtype: float64

 **************************************************

Rating missing values (%):

user_id     0.0
anime_id    0.0
rating      0.0
dtype: float64
```

Handling missing values in dataset

```
10… anime_id    0
    name        0
    genre       0
    type        0
    episodes    0
    rating      0
    members     0
    dtype: int64
```

Deleting anime with 0 rating and filling mode value for genre and type

1…

	user_id	anime_id	rating
0	1	20	NaN
1	1	24	NaN
2	1	79	NaN
3	1	226	NaN
4	1	241	NaN
5	1	355	NaN
6	1	356	NaN
7	1	442	NaN
8	1	487	NaN
9	1	846	NaN
10	1	936	NaN
11	1	1546	NaN
12	1	1692	NaN
13	1	1836	NaN
14	1	2001	NaN
15	1	2025	NaN
16	1	2144	NaN
17	1	2787	NaN
18	1	2993	NaN
19	1	3455	NaN

Filling NaN values

name	.hack//Roots	.hack//Sign	.hack//Tasogare no Udewa Densetsu	009-1	07-Ghost	11eyes	12-sai.: Chicchana Mune no Tokimeki	3 Choume no Tama: Uchi no Tama Shirimasenka?	30-sai no Hoken Taiiku	91 Days	...	Zone of the Enders: Dolores, I	Zukkoke Knight: Don De La Mancha	ef: A Tale of Melodies.	ef: A Tale of Memories.	gdgd Fairies	gdgd Fairies 2	iDOLM Xenc
user_id																		
1	NaN	NaN	NaN	NaN	NaN	6.49	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	

5 rows × 2734 columns

Pivot table for similarity

```
[18]:  anime_recommendation('Dragon Ball Z')

Recommended because you watched Dragon Ball Z:

#1: Dragon Ball, 79.32% match
#2: Fullmetal Alchemist, 42.81% match
#3: Death Note, 42.6% match
#4: Code Geass: Hangyaku no Lelouch, 37.64% match
#5: Yuu☆Yuu☆Hakusho, 37.39% match
```

Recommendation through cosine similarity for Dragon Ball Z

CONCLUSION:

Anime is a varied kind of entertainment that appeals to a wide range of audiences with a variety of genres and themes. Making personal anime recommendations becomes a difficult undertaking as a result. We have addressed this issue by creating a novel anime recommendation engine after giving it great thought. Additionally, the system has the ability to suggest random anime series for the user to check out. This concept can definitely be utilised to construct an anime recommendation system in real-time. The proposed model can be further refined as part of the future scope by basing recommendations on user age groups and utilising demographic data about the users. It would be interesting to see how these factors would affect how the recommendations varied for each user.

REFERENCES:

1.Feng, Junmei et al. (2021). “RBPR: A hybrid model for the new user cold start problem in recommender systems”. In: Knowledge-Based Systems 214, p. 106732. issn: 0950- 7051. doi: https://doi.org/10.1016/j.knosys.2020.106732.

2.Ortega, Fernando et al. (2021). “Providing reliability in recommender systems through Bernoulli Matrix Factorization”. In: Information Sciences 553, pp. 110–128. issn: 0020-0255. doi: https://doi.org/10.1016/j.ins.2020.12.001.

3.Shambour, Qusai (2021). “A deep learning based algorithm for multi-criteria recommender systems”. In: Knowledge-Based Systems 211, p. 106545. issn: 0950-7051. doi: https://doi.org/10.1016/j.knosys.2020.106545.

4. Shojaei, Mansoore and Hassan Saneifar (2021). “MFSR: A novel multi-level fuzzy similarity measure for recommender systems”. In: Expert Systems with Applications 177, p. 114969. issn: 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.114969.

5.Girsang, A S et al. (June 2020). “Collaborative Recommendation System in Users of Anime Films”. In: Journal of Physics: Conference Series 1566, p. 012057. doi: 10. 1088/1742-6596/1566/1/012057.

6.N´apoles, Gonzalo, Isel Grau and Yamisleydi Salgueiro (2020). “Recommender system using Long-term Cognitive Networks”. In: Knowledge-Based Systems 206, p. 106372. issn: 0950-7051. doi: https://doi.org/10.1016/j.knosys.2020.106372.

7.R, Kiran, Pradeep Kumar and Bharat Bhasker (2020). “DNNRec: A novel deep learning based hybrid recommender system”. In: Expert Systems with Applications 144, p. 113054. issn: 0957-4174. doi: https://doi.org/10.1016/j.eswa.2019.113054.

8.Mu, Yi et al. (2019). “An Efficient Similarity Measure for Collaborative Filtering”. In: Procedia Computer Science 147. 2018 International Conference on Identification, Information and Knowledge in the Internet of Things, pp. 416–421. issn: 1877-0509. doi: https://doi.org/10.1016/j.procs.2019.01.258.

9.Wang, Kai et al. (2019). “SDDRS: Stacked Discriminative Denoising Auto-Encoder based Recommender System”. In: Cognitive Systems Research 55, pp. 164–174. issn: 1389- 0417. doi: https://doi.org/10.1016/j.cogsys.2019.01.011.

10.Zhang, Shuai et al. (Feb. 2019). “Deep Learning Based Recommender System: A Survey and New Perspectives”. In: ACM Comput. Surv. 52.1. issn: 0360-0300. doi: 10.1145/ 3285029.

11.Nguyen, Van-Doan and Van-Nam Huynh (2017). “Two-probabilities focused combination in recommender systems”. In: International Journal of Approximate Reasoning 80, pp. 225–238. issn: 0888-613X. doi: https://doi.org/10.1016/j.ijar.2016.09. 005.

12.Nguyen, Van-Doan, Songsak Sriboonchitta and Van-Nam Huynh (2017). “Using community preference for overcoming sparsity and cold-start problems in collaborative filtering system offering soft ratings”. In: Electronic Commerce Research and Applications 26, pp. 101–108. issn: 1567-4223. doi: https://doi.org/10.1016/j.elerap. 2017.10.002.

13.Al-Badarenah A, Alsakran J (2016) An automated recommender system for course selection. International Journal of Advanced Computer Science and Applications 7(3):166–175

14.Azfar T, Haw SC (2020) Evaluation of hybrid recommender techniques on movielens dataset. PalArch's Journal of Archaeology of Egypt/Egyptology 17(10):890–902

15.Barkan O, Koenigstein N (2016) Item2vec: neural item embedding for collaborative filtering. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, pp 1–6

16.Behera RN, Saha PL, Chakraborty A, Dash S (2017) Hybrid movie recommendation system based on pso based clustering

17.Cintia Ganesha Putri D, Leu JS, Seda P (2020) Design of an unsupervised machine learning-based movie recommender system. Symmetry 12(2):185

18. Clevert DA, Unterthiner T, Hochreiter S (2015) Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:151107289

19. Geetha G, Safa M, Fancy C, Saranya D (2018) A hybrid approach using collaborative filtering and content based filtering for recommender system. In: Journal of Physics: Conference Series, IOP Publishing, vol 1000, p 012101

20.Hande R, Gutti A, Shah K, Gandhi J, Kamtikar V (2016) International journal of engineering sciences & research technology moviemender-a movie recommender system

21.Hornung T, Ziegler CN, Franz S, Przyjaciel-Zablocki M, Schätzle A, Lausen G (2013) Evaluating hybrid music recommender systems. In: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE, vol 1, pp 57–64