

Anime Recommendation Sytem

Description:

The rapid development of the Japanese animation industry has produce tons of anime movies which made interest to groups of people. Each anime movie has its own characteristic which complies with specific user's interests. Therefore, a personalization engine was needed to provide recommendations.

Problem statement:

To recommend anime to the user based on the anime he has watched.

Loading the required libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.sparse import csr_matrix
#from sklearn.utils.extmath import randomized_svd
import re
import nltk
from nltk.corpus import stopwords
from tqdm import tqdm
from nltk.stem import WordNetLemmatizer
import warnings
warnings.filterwarnings("ignore")
from nltk.stem.snowball import SnowballStemmer
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from scipy.sparse import hstack
from sklearn.metrics.pairwise import cosine_similarity
```

Dataset source: <https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020>
(<https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020>)

```
In [2]: pd.set_option('display.max_columns', None)
anime = pd.read_csv(r'D:\Anime Recommendation System\anime.csv')
anime.head(2)
```

Out[2]:

	MAL_ID	Name	Score	Genres	English name	Japanese name	Type	Episodes	Aired	Pre
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	Cowboy Bebop	カウボーイビバップ	TV	26	Apr 3, 1998 to Apr 24, 1999	Spr 199
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	Cowboy Bebop:The Movie	カウボーイビバップ 天国の扉	Movie	1	Sep 1, 2001	Un

```
In [3]: anime[anime['Score']=='Unknown'].head(2)
```

Out[3]:

	MAL_ID	Name	Score	Genres	English name	Japanese name	Type	Episodes	Air
1404	1547	Obake no Q-tarou	Unknown	Comedy, School, Slice of Life, Supernatural	Unknown	オバケのQ太郎	TV	96	Au 29, 1991 to Jun 28, 1991
1504	1656	PostPet Momobin	Unknown	Comedy, Kids	Unknown	ポストペットモモ便	TV	24	Jul 20, 2001 to Jan 9, 2002

```
In [4]: anime_list = pd.read_csv(r"D:\Anime Recommendation System\animelist.csv")
anime_list.head(2)
```

Out[4]:

	user_id	anime_id	rating	watching_status	watched_episodes
0	0	67	9	1	1
1	0	6702	7	1	4

```
In [5]: anime_with_synopsis = pd.read_csv(r"D:\Anime Recommendation System\anime_with_
synopsis.csv")
anime_with_synopsis.head(2)
```

Out[5]:

	MAL_ID	Name	Score	Genres	sypnopsis
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	In the year 2071, humanity has colonized sever...
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	other day, another bounty —such is the life of ...

Exploratory Data Analysis and Data Cleaning

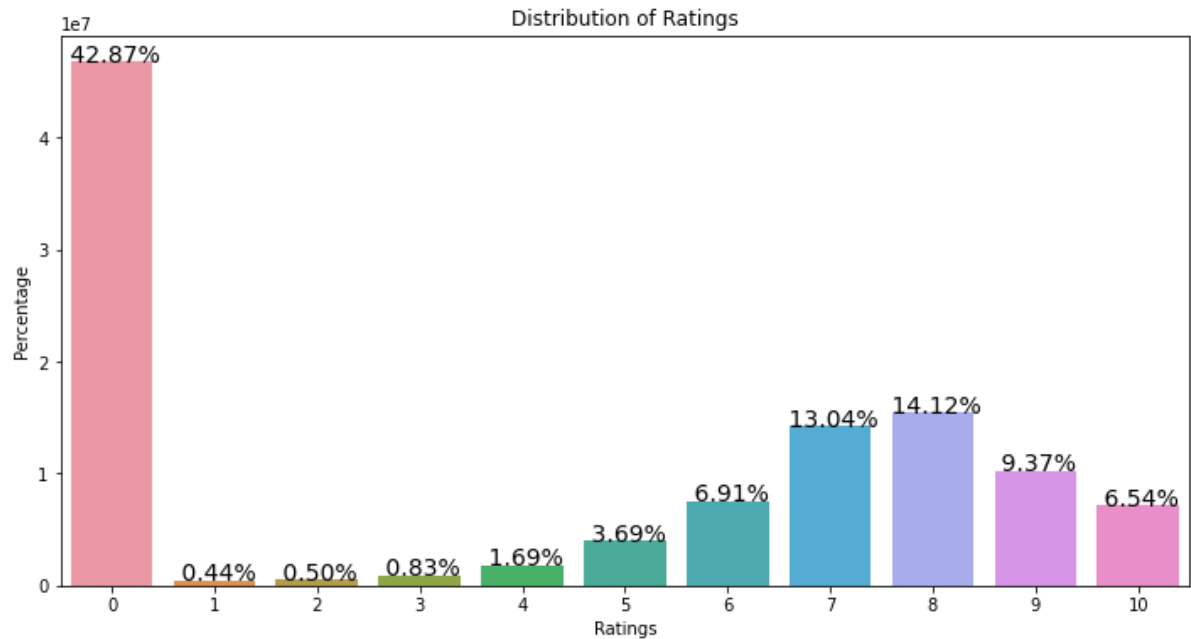
```
In [6]: def countplot_with_percentages(serie):
        ax = sns.countplot(serie)
        total_count = serie.count()

        for p in ax.patches:
            x = p.get_bbox().get_points()[:, 0]
            y = p.get_bbox().get_points()[1, 1]
            percentage = p.get_height() / total_count * 100
            ax.annotate(f'\n{percentage: .2f}%', (x.mean(), y), ha='center', size=
14)

plt.figure(figsize=(12, 6))
#plt.axes(yscale="log")

countplot_with_percentages(anime_list.rating)

plt.title("Distribution of Ratings")
plt.ylabel("Percentage")
plt.xlabel("Ratings");
```



Content Based

Data Processing

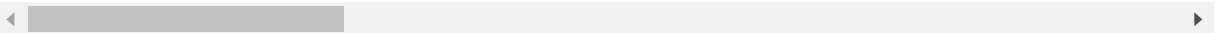
```
In [7]: anime.columns
```

```
Out[7]: Index(['MAL_ID', 'Name', 'Score', 'Genres', 'English name', 'Japanese name',  
             'Type', 'Episodes', 'Aired', 'Premiered', 'Producers', 'Licensors',  
             'Studios', 'Source', 'Duration', 'Rating', 'Ranked', 'Popularity',  
             'Members', 'Favorites', 'Watching', 'Completed', 'On-Hold', 'Dropped',  
             'Plan to Watch', 'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-  
             6',  
             'Score-5', 'Score-4', 'Score-3', 'Score-2', 'Score-1'],  
            dtype='object')
```

```
In [8]: anime_df = anime[['MAL_ID', 'Name', 'Score', 'Genres', 'English name',  
                        'Type', 'Episodes', 'Aired', 'Premiered', 'Producers', 'Licensors',  
                        'Studios', 'Source', 'Duration', 'Rating', 'Ranked', 'Popularity',  
                        'Members', 'Favorites', 'Watching', 'Completed', 'On-Hold', 'Dropped',  
                        'Plan to Watch', 'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-6',  
                        'Score-5', 'Score-4', 'Score-3', 'Score-2', 'Score-1']].copy()  
anime_df.head(2)
```

```
Out[8]:
```

	MAL_ID	Name	Score	Genres	English name	Type	Episodes	Aired	Premiered	Pr
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	Cowboy Bebop	TV	26	Apr 3, 1998 to Apr 24, 1999	Spring 1998	Ba Vi:
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	Cowboy Bebop:The Movie	Movie	1	Sep 1, 2001	Unknown	St Ba Vi:



```
In [9]: anime_df = anime_df.merge(anime_with_synopsis, how='left', on=['MAL_ID', 'Name', 'Score', 'Genres'])  
anime_df.shape
```

```
Out[9]: (17562, 35)
```

```
In [10]: anime_df.head(2)
```

Out[10]:

	MAL_ID	Name	Score	Genres	English name	Type	Episodes	Aired	Premiered	Pr
0	1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space	Cowboy Bebop	TV	26	Apr 3, 1998 to Apr 24, 1999	Spring 1998	Ba Vi:
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space	Cowboy Bebop:The Movie	Movie	1	Sep 1, 2001	Unknown	St Ba Vi:

Data Cleaning and preprocessing

Checking duplicates

```
In [11]: anime_df[anime_df.duplicated()==True]
```

Out[11]:

	MAL_ID	Name	Score	Genres	English name	Type	Episodes	Aired	Premiered	Producers
--	--------	------	-------	--------	--------------	------	----------	-------	-----------	-----------

```
In [12]: anime_df.Rating.unique()
```

```
Out[12]: array(['R - 17+ (violence & profanity)', 'PG-13 - Teens 13 or older',  
                'PG - Children', 'R+ - Mild Nudity', 'G - All Ages', 'Rx - Hentai',  
                'Unknown'], dtype=object)
```

```

In [13]: anime_df["Genres"] = anime_df["Genres"].apply(lambda x: x.replace(" ", ""))
anime_df["Genres"] = anime_df["Genres"].apply(lambda x: x.replace("-", ""))

anime_df["Producers"] = anime_df["Producers"].apply(lambda x: x.replace(" ", ""))

anime_df["Licensors"] = anime_df["Licensors"].apply(lambda x: x.replace(" ", ""))

anime_df["Studios"] = anime_df["Studios"].apply(lambda x: x.replace(" ", ""))

anime_df["Premiered"] = anime_df["Premiered"].apply(lambda x: x.replace(" ", ""))

anime_df["Rating"] = anime_df["Rating"].apply(lambda x: re.sub("[^a-zA-Z0-9\n]", " ", x))
anime_df["Rating"] = anime_df["Rating"].apply(lambda x: x.replace(" ", ""))

anime_df["synopsis"].fillna('', inplace=True)

```

```

In [14]: anime_df.head(2)

```

Out[14]:

	MAL_ID	Name	Score	Genres	English name	Type
0	1	Cowboy Bebop	8.78	Action,Adventure,Comedy,Drama,SciFi,Space	Cowboy Bebop	TV
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action,Drama,Mystery,SciFi,Space	Cowboy Bebop:The Movie	Movie

Concatenating all the required textual fields in a single column

```
In [15]: anime_df["tags"] = anime_df["Genres"] + " " + anime_df["Type"] + " " + anime_df["Premiered"] + " " + anime_df["Producers"] + " " + anime_df["Licensors"] + " " + anime_df["Studios"] + " " + anime_df["Source"] + " " + anime_df["Rating"] + " " + anime_df["synopsis"]
anime_df.head(2)
```

Out[15]:

	MAL_ID	Name	Score	Genres	English name	Type
0	1	Cowboy Bebop	8.78	Action,Adventure,Comedy,Drama,SciFi,Space	Cowboy Bebop	TV
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Action,Drama,Mystery,SciFi,Space	Cowboy Bebop:The Movie	Movie

```
In [16]: anime_df.columns
```

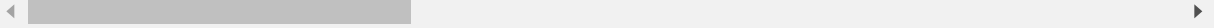
```
Out[16]: Index(['MAL_ID', 'Name', 'Score', 'Genres', 'English name', 'Type', 'Episode s',
               'Aired', 'Premiered', 'Producers', 'Licensors', 'Studios', 'Source',
               'Duration', 'Rating', 'Ranked', 'Popularity', 'Members', 'Favorites',
               'Watching', 'Completed', 'On-Hold', 'Dropped', 'Plan to Watch',
               'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-6', 'Score-5',
               'Score-4', 'Score-3', 'Score-2', 'Score-1', 'synopsis', 'tags'],
              dtype='object')
```



```
In [17]: anime_data = anime_df[['MAL_ID', 'Name', 'Score', 'English name', 'Episodes',
    'Ranked', 'Popularity', 'Members', 'Favorites',
    'Watching', 'Completed', 'On-Hold', 'Dropped', 'Plan to Watch',
    'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-6', 'Score-5',
    'Score-4', 'Score-3', 'Score-2', 'Score-1', 'tags']]
anime_data.head(2)
```

Out[17]:

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity	Members	Favorites
0	1	Cowboy Bebop	8.78	Cowboy Bebop	26	28.0	39	1251960	61971
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Cowboy Bebop:The Movie	1	159.0	518	273145	1174



```
In [18]: def decontracted(phrase):
    phrase = re.sub(r"won't", "will not", phrase)
    phrase = re.sub(r"can't", "can not", phrase)
    phrase = re.sub(r"n't", " not", phrase)
    phrase = re.sub(r"\'re", " are", phrase)
    phrase = re.sub(r"\'s", " is", phrase)
    phrase = re.sub(r"\'d", " would", phrase)
    phrase = re.sub(r"\'ll", " will", phrase)
    phrase = re.sub(r"\'t", " not", phrase)
    phrase = re.sub(r"\'ve", " have", phrase)
    phrase = re.sub(r"\'m", " am", phrase)
    return phrase
```

```
In [19]: nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\anike\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Out[19]: True

```
In [20]: stop_words = set(stopwords.words('english'))
print(stop_words)
```

```
{'by', 'down', 'to', 'that', 'until', 'herself', 'no', 'against', 'hadn't',
'needn't', 'once', 'shan', 'before', 'into', 'them', 'further', 'does', 'wh
o', 'all', 'having', 'which', 'when', 'll', 'off', 'haven't', 'being', 'would
n't', 'should've', 'whom', 'how', 'more', 't', 'me', 'you'd', 'with', 'could
n', 'mightn't', 'needn', 'am', 'such', 'should', 'weren't', 'above', 'if', 't
hrough', 'some', 'very', 'd', 'shan't', 'are', 'than', 'what', 'm', 'mightn',
'most', 'has', 'shouldn't', 'same', 'haven', 'during', 'each', 've', 'yoursel
ves', 'few', 'nor', 'under', 'mustn', 'won', 'as', 'for', 're', 'we', 'my',
'hers', 'his', 'you're', 'they', 'here', 'just', 'between', 'that'll', 'i
t's', 'because', 'ours', 'doing', 'had', 'doesn't', 'wasn't', 'ourselves', 'y
ours', 'the', 'not', 'don't', 'don', 'mustn't', 'up', 'our', 'won't', 'can',
'too', 'he', 'now', 'in', 'wouldn', 'itself', 'own', 'shouldn', 'after', 'b
e', 'you', 'she's', 'him', 'at', 'isn', 'other', 'did', 'you'll', 'her', 'doe
sn', 'ain', 'hasn't', 'those', 'again', 'hadn', 'y', 'on', 'about', 'theirs',
'hasn', 'of', 'there', 'or', 'this', 'their', 'then', 'these', 'o', 'your',
'any', 'weren', 'do', 'will', 'you've', 'wasn', 'themselves', 'a', 'and',
'i', 'an', 'aren', 'aren't', 'ma', 'while', 'from', 'so', 'but', 'been', 'it
s', 'didn', 'was', 'have', 'below', 'both', 'isn't', 'himself', 'she', 'did
n't', 'yourself', 's', 'out', 'is', 'myself', 'where', 'couldn't', 'over', 'i
t', 'were', 'why', 'only'}
```

```
In [21]: nltk.download('wordnet')
lemmatizer = WordNetLemmatizer()
```

```
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\anike\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
In [22]: snow_stemmer = SnowballStemmer(language='english')
```

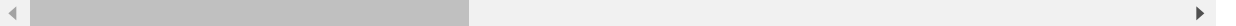
```
In [23]: #for sentence in tqdm(anime_data["tags"].values):
def text_preprocessing(text):
    text = ' '.join(words.lower() for words in text.split(" ")) if words not in
stop_words)
    text = re.sub('[^a-zA-Z0-9\n]', ' ', text)
    text = re.sub('\s+', ' ', text)
    text = re.sub("\n", " ", text)
    text = re.sub("\t", " ", text)
    text = re.sub(",", " ", text)
    text = decontracted(text)
    text = re.sub(r'[\w\s]', '', text)
    word_list = nltk.word_tokenize(text)
    #text = ' '.join(Lemmatizer.Lemmatize(word) for word in word_list)
    text = ' '.join(snow_stemmer.stem(word) for word in word_list)
    return text.strip()
```

```
In [24]: anime_data["tags"] = anime_data["tags"].apply(text_preprocessing)
```

```
In [25]: anime_data.head(2)
```

```
Out[25]:
```

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity	Members	Favorites
0	1	Cowboy Bebop	8.78	Cowboy Bebop	26	28.0	39	1251960	61971
1	5	Cowboy Bebop: Tengoku no Tobira	8.39	Cowboy Bebop:The Movie	1	159.0	518	273145	1174



```
In [26]: anime_data["tags"][0]
```

```
Out[26]: 'action adventur comedi drama scifi space tv spring1998 bandaivisu funim band  
aientertain sunris origin r17violenceprofan in year 2071 human colon sever pl  
anet moon solar system leav uninhabit surfac planet earth behind the inter so  
lar system polic attempt keep peac galaxi aid part outlaw bounti hunter refer  
cowboy the ragtag team aboard spaceship bebop two individu mellow carefre spi  
ke spiegel balanc boister pragmat partner jet black pair make live chase boun  
ti collect reward thrown cours addit new member meet travel ein genet engin h  
igh intellig welsh corgi femm fatal fay valentin enigmat trickster memori los  
s strang comput whiz kid edward wong the crew embark thrill adventur unravel  
member s dark mysteri past littl littl well balanc high densiti action light  
heart comedi cowboy bebop space western classic homag smooth improvis music n  
ame after'
```

```
In [27]: anime_data[anime_data["Score"] == "Unknown"]
```

Out[27]:

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
1404	1547	Obake no Q-tarou	Unknown	Unknown	96	11270.0	11870
1504	1656	PostPet Momobin	Unknown	Unknown	24	11598.0	12929
1578	1739	Shibawanko no Wa no Kokoro	Unknown	Unknown	80	11947.0	11799
1699	1863	Silk Road Shounen Yuuto	Unknown	Unknown	26	12096.0	12286
1893	2073	Hengen Taima Yakou Karura Mau! Sendai Kokeshi ...	Unknown	Unknown	6	14363.0	11317
1990	2173	Kouchuu Ouja Mushiking Super Battle Movie: Yam...	Unknown	Mushiking:The King of Beetles SUPER BATTLE MOV...	1	15166.0	12530
2165	2366	Touma Kishinden Oni	Unknown	Unknown	25	12591.0	11179
2422	2636	Dorami-chan: Wow, The Kid Gang of Bandits	Unknown	Unknown	1	13835.0	13232
2425	2639	Dorami-chan: Hello Kyouryuu Kids!!	Unknown	Unknown	1	13844.0	13242
2428	2642	Dorami-chan: A Blue Straw Hat	Unknown	Unknown	1	13845.0	13790

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
2475	2691	Gensei Shugoshin P-hyoro Ikka	Unknown	Unknown	1	14116.0	13295
2485	2701	Susie-chan to Marvy	Unknown	Little Susie and Marvy	104	12301.0	14538
2492	2708	Wankorobee	Unknown	Unknown	26	12782.0	14602
2495	2712	Shiratori Reiko de Gozaimasu!	Unknown	I am THE Shiratori Reiko!	1	11996.0	12892
2517	2735	Mugen Senki Portriss	Unknown	Unknown	52	15698.0	12851
2520	2738	Kasumin 3rd Season	Unknown	Unknown	26	14908.0	12369
2521	2739	Highschool Aurabuster: Hikari no Mezame	Unknown	Unknown	3	14409.0	11791
2523	2741	Monkey Turn V	Unknown	Unknown	25	15646.0	11845
2524	2742	Omoikkiri Kagaku Adventure Sou Nanda!	Unknown	Full-Blast Science Adventure: So That's How It Is!	26	11329.0	12746
2532	2750	Heisei Inu Monogatari Bow Movie	Unknown	Heisei Dog Stories: Bow	1	14333.0	13734
2676	2930	Tensai? Dr. Hamax	Unknown	Unknown	12	12443.0	14501
2795	3073	Robby to Kerobby	Unknown	Robby & Kerobby	52	11725.0	14482

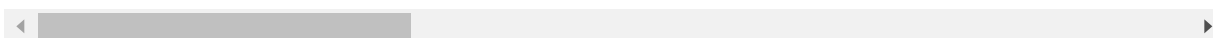
	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
2857	3164	Ichigo Ichie: Koibana Tomobana	Unknown	Unknown	13	14573.0	12626
2866	3180	Manga Aesop Monogatari (TV)	Unknown	Unknown	52	15417.0	12900
2903	3243	Elec-king The Animation	Unknown	Unknown	13	13894.0	14546
2908	3251	Hatara Kids Mai Ham Gumi	Unknown	Master Hamsters	50	14292.0	12679
2936	3290	Vampiyan Kids	Unknown	Vampiyan Kids	23	12757.0	11023
2981	3362	Oshare Majo Love and Berry: Shiawase no Mahou	Unknown	Fashionable Witches Love and Berry:Magic of Ha...	1	11403.0	11498
2988	3372	RGB Adventure	Unknown	RGB Adventure	6	11691.0	14288
3016	3414	Hula Kappa	Unknown	Unknown	78	14511.0	13636
...
17531	48365	Youkai Watch ♪	Unknown	Unknown	Unknown	Unknown	15988
17532	48369	Wo Shi Da Shenxian 2nd Season	Unknown	Da Shen Xian	16	12882.0	16796
17533	48374	Fei Ren Zai 4th Season	Unknown	Unknown	12	13980.0	16948

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
17534	48375	Mahouka Koukou no Rettousei: Tsuioku-hen	Unknown	Unknown	Unknown	Unknown	17526
17535	48391	Mazica Party	Unknown	Unknown	Unknown	Unknown	17471
17536	48405	Totsukuni no Shoujo OVA	Unknown	Unknown	1	Unknown	8462
17537	48406	Re-Main	Unknown	Unknown	Unknown	Unknown	17528
17538	48409	Chocomatsu- san: White Day-hen	Unknown	Unknown	3	Unknown	12349
17539	48411	Aria the Benedizione	Unknown	Unknown	1	Unknown	17530
17540	48413	Hataraku Maou-sama! 2nd Season	Unknown	Unknown	Unknown	Unknown	1833
17541	48414	Sabikui Bisco	Unknown	Unknown	Unknown	Unknown	10066
17542	48415	Shijou Saikyou no Daimaou, Murabito A ni Tense...	Unknown	Unknown	Unknown	Unknown	11443
17543	48417	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Unknown	Unknown	Unknown	Unknown	2597

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
17544	48418	Maou Gakuin no Futekigousha: Shijou Saikyou no...	Unknown	Unknown	Unknown	Unknown	5731
17545	48422	Kazoku: Haha to Shimai no Kyousei	Unknown	Unknown	Unknown	Unknown	15694
17546	48425	Akuma-kun (Shin Anime)	Unknown	Unknown	Unknown	Unknown	14015
17547	48426	Kitarou Tanjou: Gegege no Nazo	Unknown	Unknown	1	Unknown	10686
17548	48427	The Sun, Moon and Stars	Unknown	Unknown	1	0.0	0
17549	48438	Mahoutsukai no Yome: Nishi no Shounen to Seira...	Unknown	Unknown	3	Unknown	4546
17550	48441	Eiyuu Densetsu: Sen no Kiseki	Unknown	Unknown	Unknown	Unknown	14139
17551	48442	Shikaru Neko	Unknown	Unknown	Unknown	Unknown	16850
17553	48466	Kyoukai Senki	Unknown	Unknown	Unknown	Unknown	17561
17554	48470	D_Cide Traumerei	Unknown	Unknown	Unknown	Unknown	17557
17555	48471	Tsuki to Laika to Nosferatu	Unknown	Unknown	Unknown	Unknown	17564

	MAL_ID	Name	Score	English name	Episodes	Ranked	Popularity
17556	48480	Wan Jie Shen Zhu 3rd Season	Unknown	Unknown	48	12855.0	16457
17557	48481	Daomu Biji Zhi Qinling Shen Shu	Unknown	Unknown	Unknown	Unknown	13116
17558	48483	Mieruko-chan	Unknown	Unknown	Unknown	Unknown	17562
17559	48488	Higurashi no Naku Koro ni Sotsu	Unknown	Higurashi:When They Cry – SOTSU	Unknown	Unknown	17558
17560	48491	Yama no Susume: Next Summit	Unknown	Unknown	Unknown	Unknown	17565
17561	48492	Scarlet Nexus	Unknown	Unknown	Unknown	Unknown	17563

5141 rows × 25 columns



```
In [28]: anime_data["Score"] = anime_data["Score"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Episodes"] = anime_data["Episodes"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Ranked"] = anime_data["Ranked"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-10"] = anime_data["Score-10"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-9"] = anime_data["Score-9"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-8"] = anime_data["Score-8"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-7"] = anime_data["Score-7"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-6"] = anime_data["Score-6"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-5"] = anime_data["Score-5"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-4"] = anime_data["Score-4"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-3"] = anime_data["Score-3"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-2"] = anime_data["Score-2"].apply(lambda x: re.sub("Unknown", "-1", x))

anime_data["Score-1"] = anime_data["Score-1"].apply(lambda x: re.sub("Unknown", "-1", x))
```

TF-IDF vectorization

```
In [29]: tfidf_vectorizer = TfidfVectorizer(min_df=10, ngram_range=(1,3), max_features=5000)
anime_data_tags_tfidf = tfidf_vectorizer.fit_transform(anime_data["tags"]).toarray()
```

```
In [30]: anime_data_tags_tfidf.shape
```

```
Out[30]: (17562, 5000)
```

```
In [31]: print(anime_data_tags_tfidf[0])
```

```
[0. 0. 0. ... 0. 0. 0.]
```

```
In [32]: print("The top 200 Feature Names are: \n",tfidf_vectorizer.get_feature_names()  
[0:200])
```

The top 200 Feature Names are:

```
['000', '10', '10 year', '100', '10th', '11', '12', '12 year', '12 year ol  
d', '13', '13th', '14', '14 year', '14 year old', '15', '15 year', '15 year o  
ld', '16', '16 year', '16 year old', '17', '17 year', '17 year old', '18', '1  
9', '1pictur', '1pictur game', '1pictur light', '1pictur light novel', '1pict  
ur manga', '1pictur manga pg13teens13orold', '1pictur origin', '1st', '20',  
'20 year', '2010', '2011', '2012', '2014', '2015', '2016', '2017', '2018', '2  
019', '2020', '20th', '21', '21st', '21st centuri', '22', '24', '25', '26',  
'27', '2nd', '2nd season', '30', '300', '3d', '3rd', '4kidsentertain', '4th',  
'50', '5th', '6th', '7th', '8bit', 'abandon', 'abcanim', 'abduct', 'abil', 'a  
bl', 'aboard', 'about', 'abrupt', 'absolut', 'abus', 'academi', 'accept', 'ac  
cid', 'accident', 'accompani', 'accomplish', 'accord', 'account', 'ace', 'ach  
iev', 'acquaint', 'acquir', 'across', 'act', 'acta', 'action', 'action advent  
ur', 'action adventur comedi', 'action adventur demon', 'action adventur dram  
a', 'action adventur fantasi', 'action adventur histor', 'action adventur mag  
ic', 'action adventur mecha', 'action adventur mysteri', 'action adventur sci  
fi', 'action comedi', 'action comedi drama', 'action comedi ecchi', 'action c  
omedi fantasi', 'action demon', 'action drama', 'action drama mecha', 'action  
ecchi', 'action fantasi', 'action fantasi magic', 'action game', 'action hist  
or', 'action horror', 'action kid', 'action magic', 'action martialart', 'act  
ion mecha', 'action mecha militari', 'action mecha scifi', 'action militari',  
'action militari scifi', 'action music', 'action mysteri', 'action scifi', 'a  
ction scifi adventur', 'action scifi comedi', 'action scifi mecha', 'action s  
cifi space', 'action scifi superpow', 'action sport', 'action supernatur', 'a  
ction superpow', 'activ', 'actor', 'actress', 'actual', 'ad', 'ad seri', 'ad  
seri yet', 'ad synopsi', 'ad titl', 'ad titl help', 'adapt', 'add', 'addit',  
'admir', 'adopt', 'ador', 'adult', 'advanc', 'advantag', 'adventur', 'adventu  
r comedi', 'adventur comedi drama', 'adventur comedi fantasi', 'adventur come  
di histor', 'adventur comedi kid', 'adventur comedi mecha', 'adventur comedi  
shounen', 'adventur comedi superpow', 'adventur demon', 'adventur drama', 'ad  
ventur drama fantasi', 'adventur fantasi', 'adventur fantasi comedi', 'advent  
ur fantasi kid', 'adventur fantasi magic', 'adventur fantasi movi', 'adventur  
fantasi ona', 'adventur fantasi tv', 'adventur histor', 'adventur kid', 'adve  
ntur kid tv', 'adventur magic', 'adventur magic fantasi', 'adventur mecha',  
'adventur mecha scifi', 'adventur movi', 'adventur movi unknown', 'adventur m  
ysteri', 'adventur mysteri comedi', 'adventur romanc', 'adventur scifi', 'adv  
entur sliceoflif', 'adventur sourc', 'adventur space', 'adventur space dram  
a', 'adventur supernatur', 'adventur tv', 'advertis', 'advfilm', 'advfilm fun  
im', 'affair', 'affect', 'affic', 'affic safeti', 'afraid']
```

```
In [33]: anime_data["Score"].reshape(-1,1).shape
```

```
Out[33]: (17562, 1)
```

```
In [34]: anime_features = pd.concat([anime_data[['Score']].copy(), pd.DataFrame(anime_d  
ata_tags_tfidf.tolist()), axis=1)
```

```
In [35]: anime_features.shape
```

```
Out[35]: (17562, 5001)
```

```
In [36]: anime_features.head(2)
```

```
Out[36]:
```

	Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1
0	8.78	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	8.39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



```
In [37]: scaler = StandardScaler()
anime_features["Score"] = scaler.fit_transform(anime_features["Score"].reshape(-1, 1))
```

```
In [38]: anime_features.head(2)
```

```
Out[38]:
```

	Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1.277573	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.166068	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



```
In [39]: cosine_sim = cosine_similarity(anime_features.values, anime_features.values)
```

```
In [40]: cosine_sim.shape
```

```
Out[40]: (17562, 17562)
```

```
In [41]: print(cosine_sim[0])
```

```
[ 1.          0.68327943  0.63987852 ... -0.65130579 -0.64619423
 -0.62719443]
```

```
In [42]: anime_index = pd.Series(anime_data.index, index=anime_data.Name)
anime_index.head(2)
```

```
Out[42]: Name
Cowboy Bebop          0
Cowboy Bebop: Tengoku no Tobira  1
dtype: int64
```

```
In [43]: def get_recommendation(anime_name, similarity = cosine_sim):  
        index = anime_index[anime_name]  
        similarity_scores = sorted(list(enumerate(similarity[index])), reverse=True,  
        key=lambda x:x[1])[0:21]  
        anime_indices = [i[0] for i in similarity_scores]  
        result = anime[['Name', 'English name', 'Genres']].iloc[anime_indices].drop(index)  
        return result
```

```
In [44]: get_recommendation("Naruto")
```

Out[44]:

	Name	English name	Genres
1574	Naruto: Shippuuden	Naruto:Shippuden	Action, Adventure, Comedy, Super Power, Martia...
7021	Naruto: Shippuuden Movie 6 - Road to Ninja	Road to Ninja:Naruto the Movie	Action, Adventure, Super Power, Martial Arts, ...
6285	Naruto: Shippuuden Movie 5 - Blood Prison	Unknown	Action, Adventure, Martial Arts, Super Power, ...
9862	Boruto: Naruto the Movie	Boruto:Naruto the Movie	Action, Comedy, Martial Arts, Shounen, Super P...
5304	Naruto: Shippuuden Movie 4 - The Lost Tower	Unknown	Action, Comedy, Martial Arts, Shounen, Super P...
2267	Naruto: Shippuuden Movie 1	Naruto:Shippuden the Movie	Action, Adventure, Comedy, Fantasy, Shounen
6875	Naruto SD: Rock Lee no Seishun Full-Power Ninden	NARUTO Spin-Off:Rock Lee & His Ninja Pals	Action, Comedy, Parody
2058	Naruto: Dai Katsugeki!! Yuki Hime Shinobu Houj...	Hidden Leaf Village Grand Sports Festival	Action, Comedy, Sports, Martial Arts, Fantasy,...
6319	Naruto: Honoo no Chuunin Shiken! Naruto vs. Ko...	Chunin Exam on Fire! and Naruto vs. Konohamaru!	Action, Adventure, Martial Arts, Shounen, Supe...
3516	Naruto: Shippuuden - Shippuu! "Konoha Gakuen" Den	Naruto Shippuden:Konoha Gakuen - Special	Comedy
7549	The Last: Naruto the Movie	The Last:Naruto the Movie	Action, Super Power, Romance, Martial Arts, Sh...
557	Naruto: Takigakure no Shitou - Ore ga Eiyuu Da...	Naruto:The Lost Story - Mission:Protect the Wa...	Action, Adventure, Comedy, Shounen, Super Power
3639	Naruto: Shippuuden Movie 2 - Kizuna	Naruto:Shippuden the Movie 2 -Bonds-	Action, Martial Arts, Shounen, Supernatural
142	Mononoke Hime	Princess Mononoke	Action, Adventure, Fantasy
11340	Boruto: Naruto the Movie - Naruto ga Hokage ni...	Unknown	Action, Comedy, Super Power, Martial Arts, Sho...
12492	Boruto: Naruto Next Generations	Boruto:Naruto Next Generations	Action, Adventure, Super Power, Martial Arts, ...
7261	Gintama': Enchousen	Gintama:Enchousen	Action, Comedy, Historical, Parody, Samurai, S...

	Name	English name	Genres
4553	Naruto: Shippuuden Movie 3 - Hi no Ishi wo Tsu...	Unknown	Action, Comedy, Martial Arts, Shounen, Super P...
14293	Gintama.: Shirogane no Tamashii-hen - Kouhan-sen	Gintama.:Silver Soul Arc - Second Half War	Action, Sci-Fi, Comedy, Historical, Parody, Sa...
6474	Hunter x Hunter (2011)	Hunter x Hunter	Action, Adventure, Fantasy, Shounen, Super Power

```
In [45]: get_recommendation("Shingeki no Kyojin")
```

Out[45]:

	Name	English name	Genres
9383	Shingeki no Kyojin Season 2	Attack on Titan Season 2	Action, Military, Mystery, Super Power, Drama,...
13252	Shingeki no Kyojin Season 3	Attack on Titan Season 3	Action, Military, Mystery, Super Power, Drama,...
15926	Shingeki no Kyojin: The Final Season	Attack on Titan Final Season	Action, Military, Mystery, Super Power, Drama,...
14963	Shingeki no Kyojin Season 3 Part 2	Attack on Titan Season 3 Part 2	Action, Drama, Fantasy, Military, Mystery, Sho...
13804	Shingeki no Kyojin Season 2 Movie: Kakusei no ...	Unknown	Action, Drama, Fantasy, Shounen, Super Power
7879	Shingeki no Kyojin OVA	Unknown	Action, Super Power, Drama, Fantasy, Shounen
64	Mobile Suit Zeta Gundam	Mobile Suit Zeta Gundam	Military, Sci-Fi, Space, Drama, Romance, Mecha
3971	Fullmetal Alchemist: Brotherhood	Fullmetal Alchemist:Brotherhood	Action, Military, Adventure, Comedy, Drama, Ma...
741	Ginga Eiyuu Densetsu	Legend of the Galactic Heroes	Military, Sci-Fi, Space, Drama
2656	Code Geass: Hangyaku no Lelouch R2	Code Geass:Lelouch of the Rebellion R2	Action, Military, Sci-Fi, Super Power, Drama, ...
16841	Shingeki no Kyojin: Chronicle	Unknown	Action, Military, Mystery, Super Power, Drama,...
9384	Shingeki no Kyojin: Kuinaki Sentaku	Unknown	Action, Fantasy, Shoujo
6474	Hunter x Hunter (2011)	Hunter x Hunter	Action, Adventure, Fantasy, Shounen, Super Power
12002	No Game No Life: Zero	Unknown	Game, Supernatural, Drama, Romance, Fantasy
14631	Kimetsu no Yaiba	Demon Slayer:Kimetsu no Yaiba	Action, Demons, Historical, Shounen, Supernatural
13293	Sayonara no Asa ni Yakusoku no Hana wo Kazarou	Maquia:When the Promised Flower Blooms	Drama, Fantasy
708	Hellsing Ultimate	Hellsing Ultimate	Action, Military, Horror, Supernatural, Vampir...
6006	Gintama'	Gintama Season 2	Action, Sci-Fi, Comedy, Historical, Parody, Sa...

	Name	English name	Genres
1822	Tengen Toppa Gurren Lagann	Gurren Lagann	Action, Adventure, Comedy, Mecha, Sci-Fi
9019	Shingeki no Kyojin Movie 2: Jiyuu no Tsubasa	Attack on Titan:Wings of Freedom	Action, Super Power, Drama, Fantasy, Shounen