

In [3]:
import pandas as pd
import numpy as np
data=pd.read_csv("D:\\IPSITA\\amazon_prime.csv")

In [4]:
data['imdb_score']=data['imdb_score'].fillna(data['imdb_score'].mean())

In [5]:
data.head(20)

	id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
0	ts20945	The Three Stooges	SHOW	The Three Stooges were an American vaudeville ...	1934	TV-PG	19	['comedy', 'family', 'animation', 'action', 'f...	['US']	26.0	tt0850645	8.6	1092.0	15.424	7.6
1	tm19248	The General	MOVIE	During America's Civil War, Union spies steal ...	1926	NaN	78	['action', 'drama', 'war', 'western', 'comedy'...	['US']	NaN	tt0017925	8.2	89766.0	8.647	8.0
2	tm82253	The Best Years of Our Lives	MOVIE	It's the hope that sustains the spirit of ever...	1946	NaN	171	['romance', 'war', 'drama']	['US']	NaN	tt0036868	8.1	63026.0	8.435	7.8
3	tm83884	His Girl Friday	MOVIE	Hildy, the journalist former wife of newspaper...	1940	NaN	92	['comedy', 'drama', 'romance']	['US']	NaN	tt0032599	7.8	57835.0	11.270	7.4
4	tm56584	In a Lonely Place	MOVIE	An aspiring actress begins to suspect that her...	1950	NaN	94	['thriller', 'drama', 'romance']	['US']	NaN	tt0042593	7.9	30924.0	8.273	7.6
5	tm160494	Stagecoach	MOVIE	A group of people traveling on a stagecoach fi...	1939	NaN	96	['western', 'drama']	['US']	NaN	tt0031971	7.8	48149.0	11.786	7.7
6	tm87233	It's a Wonderful Life	MOVIE	A holiday favourite for generations... George...	1946	PG	130	['drama', 'family', 'fantasy', 'romance', 'com...	['US']	NaN	tt0038650	8.6	444243.0	26.495	8.3
7	tm19424	Detour	MOVIE	The life of Al Roberts, a pianist in a New Yor...	1945	NaN	66	['thriller', 'drama', 'crime']	['US']	NaN	tt0037638	7.3	17233.0	7.757	7.2
8	tm116781	My Man Godfrey	MOVIE	Fifth Avenue socialite Irene Bullock needs a "...	1936	NaN	95	['comedy', 'romance', 'drama']	['US']	NaN	tt0028010	8.0	23532.0	8.633	7.6
9	tm112005	Marihuana	MOVIE	A young girl named Burma attends a beach party...	1936	NaN	57	['crime', 'drama']	['US']	NaN	tt0026683	4.0	864.0	3.748	3.6
10	tm22806	Intolerance: Love's Struggle Throughout the Ages	MOVIE	The story of a poor young woman, separated by ...	1916	NaN	197	['history', 'drama']	['US']	NaN	tt0006864	7.7	15242.0	9.412	7.2
11	tm88001	The Stranger	MOVIE	A man working for the War Crimes Commission su...	1946	NaN	95	['thriller', 'crime', 'drama']	['US']	NaN	tt0038991	7.3	26293.0	7.566	7.3
12	tm141715	The Phantom of the Opera	MOVIE	A grotesquely disfigured composer known as "Th...	1925	NaN	101	['horror']	['US']	NaN	tt0016220	7.5	18190.0	7.428	7.2
13	tm112415	Tell Your Children	MOVIE	High-school principal Dr. Alfred Carroll relat...	1938	PG	66	['drama', 'crime']	['US']	NaN	tt0028346	3.7	8617.0	5.412	4.1
14	tm2838	The Most Dangerous Game	MOVIE	When legendary hunter Bob Rainsford is shipwre...	1932	NaN	63	['thriller', 'horror', 'action']	['US']	NaN	tt0023238	7.1	12203.0	8.828	7.0
15	tm155610	Wuthering Heights	MOVIE	The Earnshaws are Yorkshire farmers during the...	1939	NaN	104	['drama', 'romance']	['US']	NaN	tt0032145	7.5	17956.0	12.336	7.3
16	tm3781	The Woman in the Window	MOVIE	A seductive woman gets an innocent professor m...	1944	NaN	99	['drama', 'thriller', 'crime']	['US']	NaN	tt0037469	7.6	15909.0	6.018	7.5
17	tm3541	Steamboat Bill, Jr.	MOVIE	The just out of college effete son of a no-non...	1928	NaN	70	['drama', 'romance', 'family', 'action', 'come...	['US']	NaN	tt0019421	7.8	14499.0	7.754	7.7
18	tm82560	The Little Foxes	MOVIE	The ruthless, moneyed Hubbard clan lives in, a...	1941	NaN	115	['romance', 'drama']	['US']	NaN	tt0033836	7.9	11775.0	5.906	7.5
19	tm64852	Stella Dallas	MOVIE	After divorcing a society man, a small-town wo...	1937	NaN	106	['drama', 'romance']	['US']	NaN	tt0029608	7.4	5440.0	5.377	6.9

In [6]:
pivot=pd.pivot_table(data,index="type",values="imdb_score",aggfunc='mean')

In [7]:
print(pivot)

	imdb_score
MOVIE	5.820003
SHOW	6.957621

In [8]:
df=pd.DataFrame(data)
new_data=df[data["genres"].str.contains("comedy")]

In [9]:
new_data.head()

	id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
0	ts20945	The Three Stooges	SHOW	The Three Stooges were an American vaudeville ...	1934	TV-PG	19	['comedy', 'family', 'animation', 'action', 'f...	['US']	26.0	tt0850645	8.6	1092.0	15.424	7.6
1	tm19248	The General	MOVIE	During America's Civil War, Union spies steal ...	1926	NaN	78	['action', 'drama', 'war', 'western', 'comedy'...	['US']	NaN	tt0017925	8.2	89766.0	8.647	8.0
3	tm83884	His Girl Friday	MOVIE	Hildy, the journalist former wife of newspaper...	1940	NaN	92	['comedy', 'drama', 'romance']	['US']	NaN	tt0032599	7.8	57835.0	11.270	7.4
6	tm87233	It's a Wonderful Life	MOVIE	A holiday favourite for generations... George...	1946	PG	130	['drama', 'family', 'fantasy', 'romance', 'com...	['US']	NaN	tt0038650	8.6	444243.0	26.495	8.3
8	tm116781	My Man Godfrey	MOVIE	Fifth Avenue socialite Irene Bullock needs a "...	1936	NaN	95	['comedy', 'romance', 'drama']	['US']	NaN	tt0028010	8.0	23532.0	8.633	7.6

In [12]:
import pandas as pd
unique_genres = set()

Iterate through each list of genres and collect unique genres
for genre_list in df['genres']:
 unique_genres.update(genre_list)

print("Unique Genres:")
for genre in unique_genres:
 print(genre)

Unique Genres:
,
u

[
i
]
a
p
f
o
e
d
n
c
,
h
r
l
w
m
y
t
s

In [13]:
data=df[data["release_year"]>=2000]

In [14]:
data.head()

	id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
2217	ts20981	Survivor	SHOW	A reality show contest where sixteen or more c...	2000	TV-PG	48	['reality']	['US']	42.0	tt0239195	7.4	24368.0	73.107	7.3
2225	tm28243	The Gift	MOVIE	Annie Wilson, young widow and mother of three,...	2000	R	112	['drama', 'thriller', 'documentation', 'action'...	['US']	NaN	tt0219699	6.7	67946.0	14.202	6.4
2226	tm23114	O Brother, Where Art Thou?	MOVIE	In the deep south during the 1930s, three esca...	2000	PG-13	107	['comedy', 'crime', 'music']	['US', 'FR', 'GB']	NaN	tt0190590	7.7	306270.0	19.293	7.3
2235	tm21856	Cast Away	MOVIE	Chuck Nolan, a top international manager for F...	2000	PG-13	143	['drama', 'romance']	['US']	NaN	tt0162222	7.8	569278.0	53.088	7.7
2236	tm18133	Unbreakable	MOVIE	An ordinary man makes an extraordinary discove...	2000	PG-13	106	['thriller', 'drama', 'scifi', 'european']	['US']	NaN	tt0217869	7.3	410474.0	30.819	7.1

In [15]:
data = data[[len(genres) > 0 for genres in data['genres']]]

In [32]:
data.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7147 entries, 2217 to 9870
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype
---  -
0   id                   7147 non-null   object
1   title                7147 non-null   object
2   type                 7147 non-null   object
3   description           7035 non-null   object
4   release_year         7147 non-null   int64
5   age_certification    2437 non-null   object
6   runtime              7147 non-null   int64
7   genres               7147 non-null   object
8   production_countries 7147 non-null   object
9   seasons              1227 non-null   float64
10  imdb_id              6570 non-null   object
11  imdb_score           7147 non-null   float64
12  imdb_votes           6238 non-null   float64
13  tmdb_popularity       6634 non-null   float64
14  tmdb_score           5498 non-null   float64
dtypes: float64(5), int64(2), object(8)
memory usage: 893.4+ KB
```

In [16]:
df.to_csv('revised_data.csv')

In []: