In [1]:
```
# importing libraries
import pandas as pd
import numpy as np
```

In [2]:
```
#Read the movie Dataset
i_cols = ['movie_id', 'movie title' ,'release date','video release date', 'IMDb
    'Animation', 'Children\'s', 'Comedy', 'Crime', 'Documentary', 'Drama', 'Fa
    'Film-Noir', 'Horror', 'Musical', 'Mystery', 'Romance', 'Sci-Fi', 'Thrille
items = pd.read_csv('ml-100k/u.item', sep='|', names=i_cols,encoding='latin-1'
```

In [3]:
```
items.head(3)
```

Out[3]:

	movie_id	movie title	release date	video release date	IMDb URL	unknown	Action	Adventure
0	1	Toy Story (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Toy%20Story%2...	0	0	0
1	2	GoldenEye (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?GoldenEye%20(...	0	1	1
2	3	Four Rooms (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Four%20Rooms%...	0	0	0

3 rows × 24 columns

In [4]:
```
# Select only needed Item
movie_content=items[['movie_id','movie title','Action', 'Adventure',
    'Animation', 'Children\'s', 'Comedy', 'Crime', 'Documentary', 'Drama', 'Fa
    'Film-Noir', 'Horror', 'Musical', 'Mystery', 'Romance', 'Sci-Fi', 'Thrille
```

In [5]:
```
movie_content.head(5)
```

Out[5]:

	movie_id	movie title	Action	Adventure	Animation	Children's	Comedy	Crime	Documentary
0	1	Toy Story (1995)	0	0	1	1	1	0	0
1	2	GoldenEye (1995)	1	1	0	0	0	0	0
2	3	Four Rooms (1995)	0	0	0	0	0	0	0
3	4	Get Shorty (1995)	1	0	0	0	1	0	0
4	5	Copycat (1995)	0	0	0	0	0	1	0

In [6]:
```
#read the ratings dataset
r_cols = ['user_id', 'movie_id', 'rating', 'unix_timestamp']
ratings = pd.read_csv('ml-100k/u.data', sep='\t', names=r_cols,encoding='latin
```

In [7]:
```
ratings
```

Out[7]:

	user_id	movie_id	rating	unix_timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596
...	...	...	...	...
99995	880	476	3	880175444
99996	716	204	5	879795543
99997	276	1090	1	874795795
99998	13	225	2	882399156
99999	12	203	3	879959583

100000 rows × 4 columns

In [8]:
```
#Added Most watched column to get most watched movies
ratings['Most Watched'] = ratings.groupby('movie_id')['movie_id'].transform('c
```

In [9]:
```
ratings
```

Out[9]:

	user_id	movie_id	rating	unix_timestamp	Most Watched
0	196	242	3	881250949	117
1	186	302	3	891717742	297
2	22	377	1	878887116	13
3	244	51	2	880606923	81
4	166	346	1	886397596	126
...	...	...	...	...	...
99995	880	476	3	880175444	160
99996	716	204	5	879795543	350
99997	276	1090	1	874795795	37
99998	13	225	2	882399156	109
99999	12	203	3	879959583	182

100000 rows × 5 columns

In [10]:
```
#Sorted Most watched by descending order
ratings = ratings.sort_values(by='Most Watched', ascending=False)
```

In [11]:
```
ratings
```

Out[11]:

	user_id	movie_id	rating	unix_timestamp	Most Watched
24954	352	50	5	884289693	583
87299	940	50	4	885921542	583
37666	517	50	5	892660727	583
88523	910	50	5	880822060	583
857	79	50	4	891271545	583
...	...	...	...	...	...
16477	76	1156	3	879576233	1
77635	60	1122	5	883326498	1
20595	405	1580	1	885549543	1
33960	409	1593	4	881108971	1
41281	195	1414	2	874825826	1

100000 rows × 5 columns

In [12]:
```
#combined both ratings and movie content datasets
combined_dataset = pd.merge(movie_content, ratings, on='movie_id')
```

In [13]:
```
combined_dataset
```

Out[13]:

	movie_id	movie title	Action	Adventure	Animation	Children's	Comedy	Crime	Documenta
0	1	Toy Story (1995)	0	0	1	1	1	0	
1	1	Toy Story (1995)	0	0	1	1	1	0	
2	1	Toy Story (1995)	0	0	1	1	1	0	
3	1	Toy Story (1995)	0	0	1	1	1	0	
4	1	Toy Story (1995)	0	0	1	1	1	0	
...	...	...	...	...	...	...	...	...	...
99995	1678	Mat' i syn (1997)	0	0	0	0	0	0	
99996	1679	B. Monkey (1998)	0	0	0	0	0	0	
99997	1680	Sliding Doors (1998)	0	0	0	0	0	0	
99998	1681	You So Crazy (1994)	0	0	0	0	1	0	
99999	1682	Scream of Stone (Schrei aus Stein) (1991)	0	0	0	0	0	0	

100000 rows × 24 columns

In [14]:
```
#Removed duplicates
combined_dataset = combined_dataset.drop_duplicates(subset='movie_id')
```

In [15]:
```
combined_dataset
```

Out[15]:

	movie_id	movie title	Action	Adventure	Animation	Children's	Comedy	Crime	Documer
0	1	Toy Story (1995)	0	0	1	1	1	0	
452	2	GoldenEye (1995)	1	1	0	0	0	0	
583	3	Four Rooms (1995)	0	0	0	0	0	0	
673	4	Get Shorty (1995)	1	0	0	0	1	0	
882	5	Copycat (1995)	0	0	0	0	0	1	
...	...	...	...	...	...	...	...	...	
99995	1678	Mat' i syn (1997)	0	0	0	0	0	0	
99996	1679	B. Monkey (1998)	0	0	0	0	0	0	
99997	1680	Sliding Doors (1998)	0	0	0	0	0	0	
99998	1681	You So Crazy (1994)	0	0	0	0	1	0	
99999	1682	Scream of Stone (Schrei aus Stein) (1991)	0	0	0	0	0	0	

1682 rows × 24 columns

In [16]:
```
combined_dataset.sort_values(by='Most Watched', ascending=False)
```

Out[16]:

	movie_id	movie title	Action	Adventure	Animation	Children's	Comedy	Crime	Documen
6402	50	Star Wars (1977)	1	1	0	0	0	0	
39421	258	Contact (1997)	0	0	0	0	0	0	
14384	100	Fargo (1996)	0	0	0	0	0	1	
26164	181	Return of the Jedi (1983)	1	1	0	0	0	0	
45747	294	Liar Liar (1997)	0	0	0	0	1	0	
...	...	...	...	...	...	...	...	...	
99801	1576	Hungarian Fairy Tale, A (1987)	0	0	0	0	0	0	
99802	1577	Death in the Garden (Mort en ce jardin, La) (1...	0	0	0	0	0	0	
98776	1348	Every Other Weekend (1990)	0	0	0	0	0	0	
99805	1579	Baton Rouge (1988)	0	0	0	0	0	0	
99999	1682	Scream of Stone (Schrei aus Stein) (1991)	0	0	0	0	0	0	

1682 rows × 24 columns

In [17]:
```
# Filter movies that belong to the 'Action' category
action_movies = combined_dataset[combined_dataset['Action'] == 1]

# Display the top 5 most watched action movies
top_5_action_movies = action_movies.head(5)
print(top_5_action_movies[['movie_id', 'movie title', 'Most Watched']])
```

```
      movie_id                    movie title  Most Watched
452          2                GoldenEye (1995)           131
673          4               Get Shorty (1995)           209
3195        17     From Dusk Till Dawn (1996)            92
3438        21  Muppet Treasure Island (1996)            84
3522        22                Braveheart (1995)           297
```

In [18]:
```
top_movies_by_category = pd.DataFrame(columns=['Action', 'Adventure',
    'Animation', 'Children\'s', 'Comedy', 'Crime', 'Documentary', 'Drama', 'Fa
    'Film-Noir', 'Horror', 'Musical', 'Mystery', 'Romance', 'Sci-Fi', 'Thrille
```

In [19]:
```
top_movies_by_category
```

Out[19]:

Action	Adventure	Animation	Children's	Comedy	Crime	Documentary	Drama	Fantasy	Film Noi

In [20]:
```
# Initialize an empty DataFrame to store top movies for all categories
top_movies_by_category = pd.DataFrame(columns=['Action', 'Adventure', 'Animati
                                               'Documentary', 'Drama', 'Fantas
                                               'Mystery', 'Romance', 'Sci-Fi',

# Loop through each category and extract top 5 movies
for column in combined_dataset.columns[2:-4]:  # Exclude 'movie_id', 'movie ti
    top_movies = combined_dataset[combined_dataset[column] == 1].nlargest(5, '
    top_movies = top_movies[['movie title']]
    top_movies_by_category[column] = top_movies['movie title'].values

# Display the resulting DataFrame
print(top_movies_by_category)
```

	Action	Adventure	Animation	Children's	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir	Horror	Musical	Mystery	Romance	Sci-Fi	Thriller
0	Star Wars (1977)	Star Wars (1977)	Toy Story (1995)	Toy Story (1995)	Liar Liar (1997)	Fargo (1996)	Hoop Dreams (1994)	Contact (1997)	E.T. the Extra-Terrestrial (1982)	L.A. Confidential (1997)	Scream (1996)	Evita (1996)	Mission: Impossible (1996)	Star Wars (1977)	Star Wars (1977)	Fargo (1996)
1	Return of the Jedi (1983)	Return of the Jedi (1983)	Lion King, The (1994)	Willy Wonka and the Chocolate Factory (1971)	Toy Story (1995)	Godfather, The (1972)	Crumb (1994)	Fargo (1996)	Nutty Professor, The (1996)	Blade Runner (1982)	Alien (1979)	Blues Brothers, The (1980)	L.A. Confidential (1997)	Return of the Jedi (1983)	Contact (1997)	Scream (1996)
2	Air Force One (1997)	Raiders of the Lost Ark (1981)	Aladdin (1992)	E.T. the Extra-Terrestrial (1982)	Back to the Future (1985)	Pulp Fiction (1994)	Celluloid Closet, The (1995)	English Patient, The (1996)	Dragonheart (1996)	Chinatown (1974)	Jaws (1975)	Wizard of Oz, The (1939)	Conspiracy Theory (1997)	English Patient, The (1996)	Return of the Jedi (1983)	Air Force One (1997)
3	Independence Day (ID4) (1996)	Rock, The (1996)	Beauty and the Beast (1991)	Wizard of Oz, The (1939)	Willy Wonka and the Chocolate Factory (1971)	L.A. Confidential (1997)	Looking for Richard (1996)	Godfather, The (1972)	Mask, The (1994)	Maltese Falcon, The (1941)	Psycho (1960)	Sound of Music, The (1965)	Amadeus (1984)	Jerry Maguire (1996)	Independence Day (ID4) (1996)	Silence of the Lambs, The (1991)
4	Raiders of the Lost Ark (1981)	Empire Strikes Back, The (1980)	Fantasia (1940)	Lion King, The (1994)	Princess Bride, The (1987)	Usual Suspects, The (1995)	Koyaanisqatsi (1983)	Pulp Fiction (1994)	Jumanji (1995)	Manchurian Candidate, The (1962)	Shining, The (1980)	Lion King, The (1994)	2001: A Space Odyssey (1968)	Empire Strikes Back, The (1980)	Twelve Monkeys (1995)	Rock, The (1996)

	War	Western
0	Star Wars (1977)	Dances with Wolves (1990)
1	Return of the Jedi (1983)	Butch Cassidy and the Sundance Kid (1969)
2	English Patient, The (1996)	Unforgiven (1992)
3	Independence Day (ID4) (1996)	Good, The Bad and The Ugly, The (1966)
4	Empire Strikes Back, The (1980)	Maverick (1994)

In [21]:
```
# Export the DataFrame to a CSV file
top_movies_by_category.to_csv('top_movies_by_category.csv', index=False)
```

In [ ]:
```

```