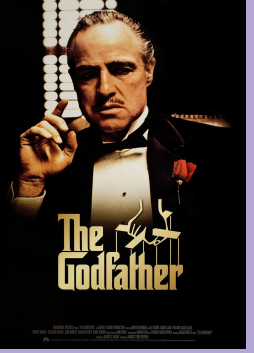


Open Analysis of the 500 Most Popular Movies

Abdul Toure

Research Question :

How does a movie's popularity and ratings influence audience preferences and reception?

Data Set Overview

- **Data Set Name :** Top 500 Popular Movies Dataset
- **Source :** Kaggle
- **Format :** CSV
- **Size :** 10KB
- **Rows :** 500
- **Columns :** 4
- **Column Names :** Title, Release Date, Popularity, Vote Average

Data Set Display

title	release_date	popularity	vote_average
Avatar: The Way of Water	2022-12-14	6789.789	7.7
Winnie the Pooh: Blood and Honey	2023-01-27	3258.54	5.9
Cocaine Bear	2023-02-22	2781.045	6.6
John Wick: Chapter 4	2023-03-22	2803.482	8.2
Puss in Boots: The Last Wish	2022-12-07	1531.792	8.3
Prizefighter: The Life of Jem Belcher	2022-06-30	1400.798	6.2
The Devil Conspiracy	2023-01-13	1280.108	6.4
Knock at the Cabin	2023-02-01	1192.041	6.4
Black Panther: Wakanda Forever	2022-11-09	1203.24	7.3
Cazadora	2023-01-19	1084.456	6.6
Plane	2023-01-12	1001.16	6.9
The Magician's Elephant	2023-03-10	923.055	7.4
Batman: The Doom That Came to Gotham	2023-03-10	978.014	6.3
John Wick: Chapter 2	2017-02-08	894.616	7.3
Money Shot: The Pornhub Story	2023-03-15	935.276	5.8
Sayen	2023-03-03	917.859	6.2
Consecration	2023-02-09	869.954	6.9
Supercell	2023-03-17	1249.672	6.1
John Wick: Chapter 5		566.145	0
Creed III	2023-03-01	805.239	7.1
Die Hart	2023-02-22	801.171	6.2
Shark Side of the Moon	2022-08-12	800.669	5.3
John Wick: Chapter 3 - Parabellum	2019-05-15	795.388	7.4
The Forbidden Legend: Sex & Chopsticks 2	2009-03-04	929.318	6
Attack on Titan	2022-09-30	1056.236	6.4
Bandit	2022-09-23	550.764	6.1
The Exorcist	2022-11-02	725.392	7
A Man Called Otto	2022-12-28	469.664	7.9
Diabolik - Ginko Attacks	2022-11-17	486.956	5.7
¡Que Viva México!	2023-03-23	393.048	4.5
Fall	2022-08-11	476.606	7.3
Demon Slayer -Kimetsu no Yaiba- The Movie: Mugen Train	2020-10-16	421.827	8.3
The Last Heretic	2022-11-09	439.846	5.3
Mummies	2023-01-05	446.228	6.6
The Whale	2022-12-09	438.185	8
Avatar	2009-12-15	432.987	7.6
Gold Run	2022-12-15	578.721	6.5
Dungeons & Dragons: Honor Among Thieves	2023-03-24	412.073	7.9
JUNG_E	2023-01-20	394.392	6.3
Scream VI	2023-03-08	400.909	7.3
The Park	2023-03-02	546.854	6.7
Troll	2022-12-01	362.028	6.8
Huesera: The Bone Woman	2023-02-10	407.004	5.8
El último hombre sobre la Tierra	2022-12-22	461.389	5.5
Prey for the Devil	2022-10-23	355.019	7.2
There Are No Saints	2022-05-27	362.223	5.3
Roald Dahl's Matilda the Musical	2022-11-25	345.559	6.8
Legion of Super-Heroes	2023-02-07	370.555	6.4
All Quiet on the Western Front	2022-10-07	347.997	7.8
The Simpsons Meet the Bocellis in Feliz Navidad	2022-12-15	355.32	5.1
Noise	2023-03-17	271.212	5
Transfusion	2023-01-05	346.454	6.5
The Help	2011-08-09	478.577	8.2
Kompromat	2022-09-07	293.321	6.4
Spider-Man: No Way Home	2021-12-15	348.363	8
Scream	2022-01-12	326.209	6.7
Terrifier 2	2022-10-06	334.634	7
H.P. Lovecraft's Witch House	2022-06-29	465.06	5.4
In His Shadow	2023-03-17	318.974	6.2
Sniper: The White Raven	2022-05-03	330.58	7.6
Boston Strangler	2023-03-17	294.656	6.8
The Enforcer	2022-09-22	330.833	7.4
Suzume	2022-11-11	329.434	8.3
Top Gun: Maverick	2022-05-24	331.46	8.3
Big Trip 2: Special Delivery	2022-10-07	313.767	6.5
Devotion	2022-11-23	300.323	7.3
R.I.P.D. 2: Rise of the Damned	2022-11-15	314.66	6.6
Thor: Love and Thunder	2022-07-06	351.829	6.6
Medieval	2022-09-08	300.925	7.2
Hex	2022-11-01	323.453	6.1
Encanto at the Hollywood Bowl	2022-12-28	330.443	7.3
Doctor Strange in the Multiverse of Madness	2022-05-04	307.946	7.4
Violent Night	2022-11-30	314.255	7.6
The Woman King	2022-09-16	283.649	7.8
Dolphin Boy	2022-04-07	272.152	6.4
Tango, Tequila and Some Lies	2023-03-09	292.267	7.6
Guillermo del Toro's Pinocchio	2022-11-09	296.664	8.3
The Chronicles of Narnia: The Lion, the Witch and the Ward	2005-12-07	287.546	7.1
Dragon Ball Super: Super Hero	2022-06-11	287.963	7.9
Four's a Crowd	2022-10-28	274.858	6.4
Prey	2022-08-02	300.661	7.8
Nocebo	2022-11-03	294.395	6.3

Analysis of Popularity Trends

- A significant number of movies exhibit a huge contrast between their popularity scores and voter ratings. Despite having high popularity, many of these films fail to get favorable reviews from audiences.

- Based on the data set, voter ratings show how satisfied the audience is. Due to promotion, a film may be seen by many people, yet if it falls short of expectations, it may still obtain low ratings. Example : Smile(2022) has a high popularity average but the voter average turned out to be a 5.4

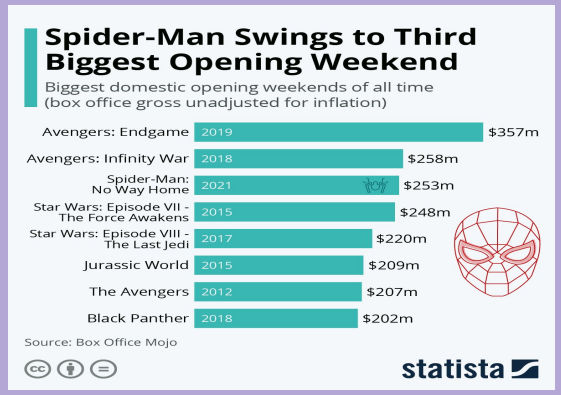
Analysis of Genre Trends

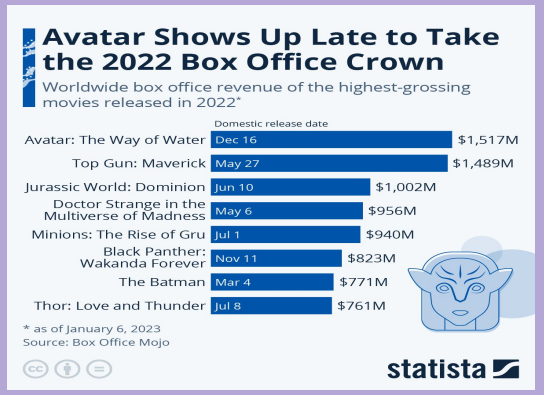
- The analysis shows that action and adventure genres tend to receive higher ratings. In contrast, horror and comedy often struggles with ratings despite sometimes being popular.

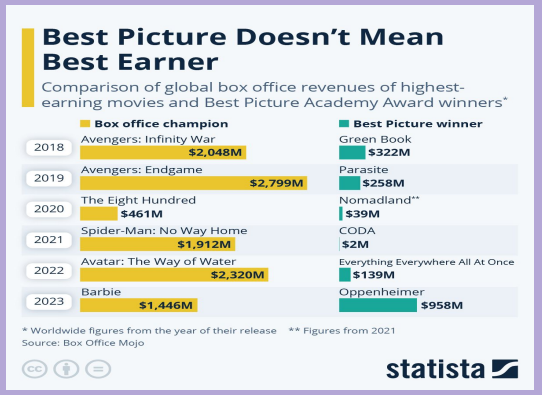
- Action movies typically have big budgets, star-studded casts, and extensive marketing campaigns, which contribute to higher popularity scores. This genre attracts a much bigger and diverse audiences while also generating a huge box office

- Despite some horror films becoming cultural icons and achieving significant box office success, they often receive lower average ratings. Horror movies are subjective so basically what scares one viewer might not scare another

Visualization used: Bar Graph







- Superhero, action, and adventure films often lead the box office but rarely win major awards like Best Picture. Popularity doesn't guarantee critical acclaim, as drama and biopic films, with stronger storytelling, tend to win more awards. It's even spotted in the datasets but instead of awards its voter averages.

Limitations In The Dataset

- **Lack of Global Context:** Without international box office data, it's challenging to assess how a movie performed in markets outside the U.S. This is particularly important for films that may perform exceptionally well in certain regions but have lower popularity or ratings elsewhere.

Techniques Used

Data Importing

```
[1]: import pandas as pd
[2]: df = pd.read_csv('top_popular_movies.csv')
[3]: df
```

Essential for manipulating and analyzing the data efficiently

Extracting Column Names

```
[10]: df.columns.tolist()
[10]: ['title', 'release_date', 'popularity', 'vote_average']
```

Helps retrieve the names of all columns and changes it to a list format

Duplicate Detection

```
duplicates_subset = df.duplicated(subset=['title'])
num_duplicates_subset = duplicates_subset.sum()
umber of duplicate rows based on 'title': {num_duplicates_subset}")

Number of duplicate rows based on 'title': 3
```

Identifies and removes duplicates which leads to better results

Subset Creation

```
umber of rows: {num_rows}, Cleaned number of rows: {new_num_rows}")

Original number of rows: 500, Cleaned number of rows: 497
```

Focuses on a subset of data that meets a certain criteria

Dataframe Sample

```
print("\nRandom 10 rows:\n", df.sample(10))
```

Random 10 rows:

	title	release_date	popularity	vote_average
339	Twilight	2008-11-20	105.	6.3
373				
340	Dog	2022-02-17	117.	7.4
585				
431	Ratatouille	2007-06-28	102.	7.8
363				
4	Puss in Boots: The Last Wish	2022-12-07	1531.	8.3
792				
491	Scream 3	2000-02-03	75.	6.0
129				
158	Lupin The 3rd vs. Cat's Eye	2023-01-26	197.	6.0
557				
164	Naked Rashomon	1973-07-04	156.	6.7
342				
278	The Maze Runner	2014-09-10	121.	7.2
824				
173	Harry Potter and the Chamber of Secrets	2002-11-13	169.	7.7
272				
5	Prizefighter: The Life of Jem Belcher	2022-06-30	1400.	6.2
798				

→

Helps better understand a variety of data points which can inform further analysis.

CONCLUSION

A movies popularity and genre strongly influence audience preferences. High ratings and critical admiration boost a movies appeal which can guide audiences toward high quality movies.