

Movie Data Analysis - PROJECT REPORT

NETFLIX



Movie Data Analysis - PROJECT REPORT



In 2007, Netflix introduced streaming services, allowing subscribers to watch movies and TV shows instantly online. This strategic shift capitalized on the growing internet bandwidth and changing consumer preferences, propelling Netflix into a leading global streaming platform. By 2010, the company began its international expansion, starting with Canada, and by 2016, it was available in over 190 countries.

Movie Data Analysis - PROJECT REPORT

As of 2024, Netflix reported a revenue of nearly \$10 billion in the third quarter, with profits reaching \$2.4 billion. The platform boasts over 283 million paid memberships across more than 190 countries, offering a vast library of TV series, films, and games in various genres and languages.

Overall, Netflix's evolution from a DVD rental service to a global streaming giant underscores its adaptability and innovative approach in the entertainment industry.



Movie Data Analysis - PROJECT REPORT

NETFLIX

Expands to 190 Countries
Around the World

REGIONS WHERE NETFLIX IS AVAILABLE



Movie Data Analysis - PROJECT REPORT

Netflix is known for its work in data science, AI, and ML, particularly for building strong recommendation models and algorithms that understand customer behavior and patterns. Suppose you are working in a data-driven job role, and you have a dataset of more than 9,000 movies. You need to solve the following questions to help the company make informed business decisions accordingly.

- 1) What is the most frequent genre of movies released on Netflix?
- 2) Which has highest votes in vote avg column?
- 3) What movie got the highest popularity? what's its genre?
- 4) What movie got the lowest popularity? what's its genre?
- 5) Which year has the most filmed movies?

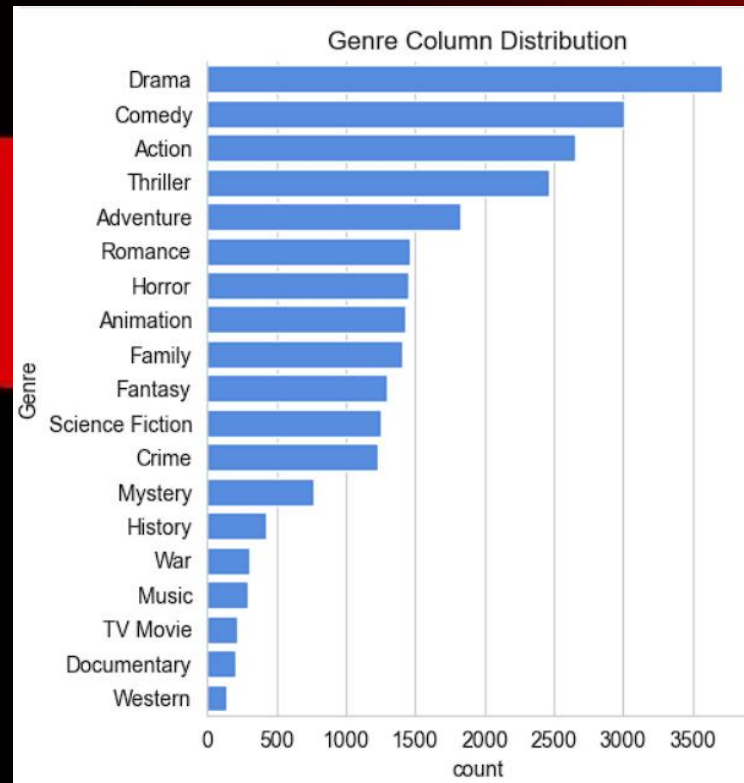
What is the most frequent genre of movies released on Netflix?

```
In [24]: df['Genre'].describe()
```

```
Out[24]: count      25551  
unique         19  
top           Drama  
freq          3715  
Name: Genre, dtype: object
```

```
In [27]: import matplotlib.pyplot as plt
```

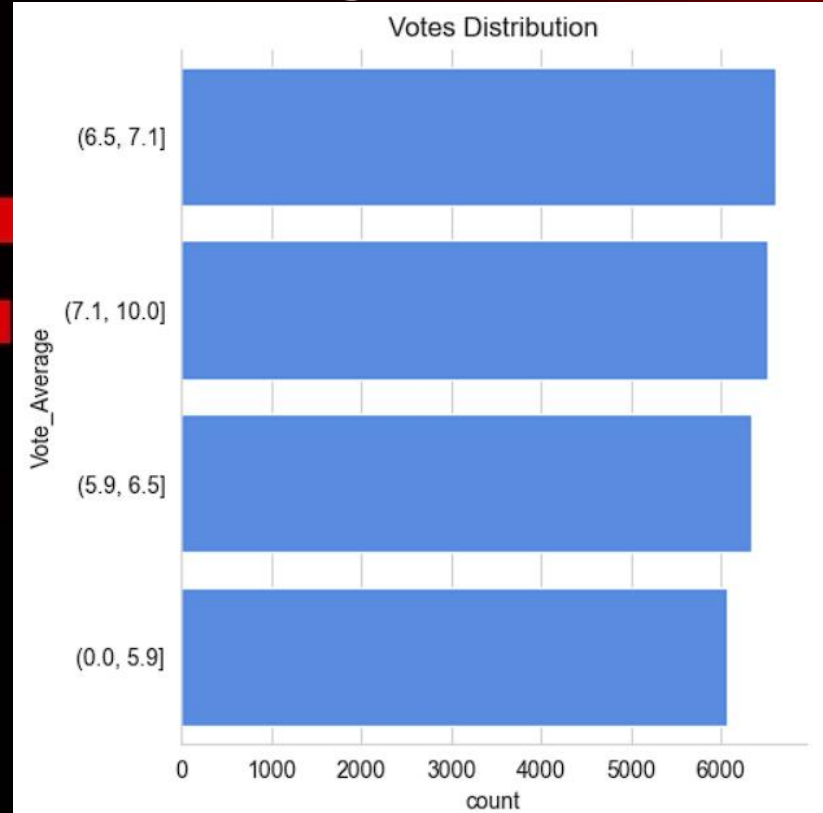
```
In [28]: sns.catplot(y= 'Genre', data= df, kind= 'count',  
                    order= df['Genre'].value_counts().index,  
                    color= '#4287f5')  
plt.title('Genre Column Distribution')  
plt.show()
```



Which has highest votes in vote avg column?

In [30]:

```
sns.catplot(y= 'Vote_Average', data= df, kind= 'count',  
            order= df['Vote_Average'].value_counts().index,  
            color= '#4287f6')  
plt.title('Votes Distribution')  
plt.show()
```



What movie got the highest popularity? what's its genre?

```
In [32]: df[df['Popularity'] == df['Popularity'].max()]
```

```
Out[32]:
```

	Release_Date	Title	Overview	Popularity	Vote_Count	Vote_Average	Original_Language	Genre
0	2021-12-15	Spider-Man: No Way Home	Peter Parker is unmasked and no longer able to...	5083.954	8940	(7.1, 10.0]	en	Action
1	2021-12-15	Spider-Man: No Way Home	Peter Parker is unmasked and no longer able to...	5083.954	8940	(7.1, 10.0]	en	Adventure
2	2021-12-15	Spider-Man: No Way Home	Peter Parker is unmasked and no longer able to...	5083.954	8940	(7.1, 10.0]	en	Science Fiction

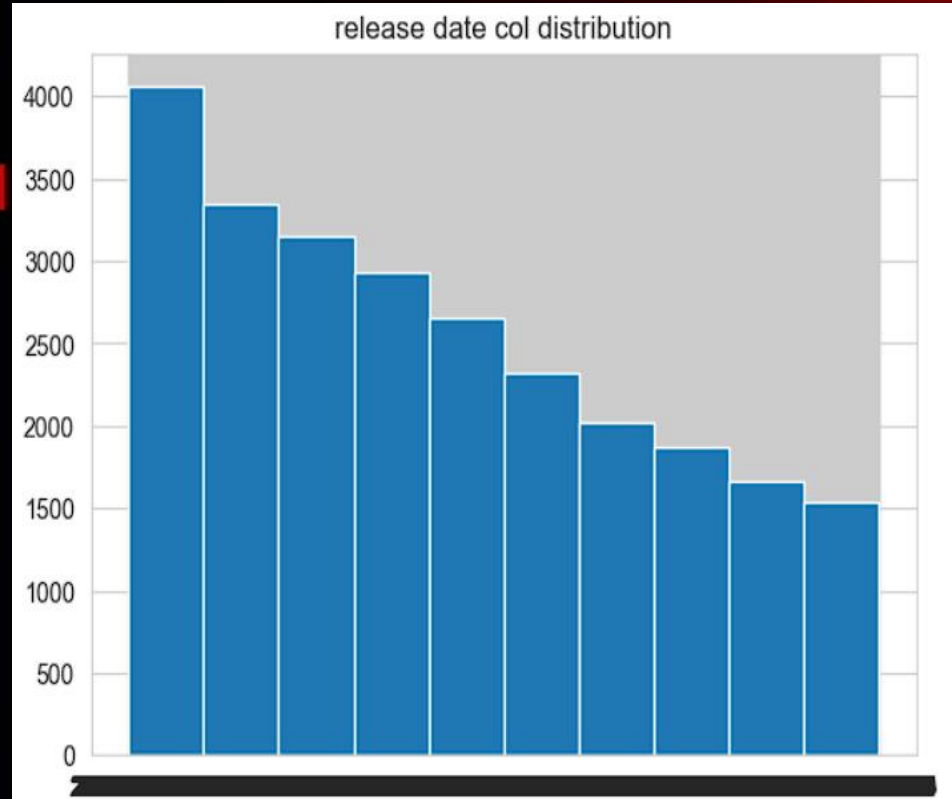
What movie got the lowest popularity? what's its genre?

```
In [33]: df[df['Popularity'] == df['Popularity'].min()]
```

Out[33]:	Release_Date	Title	Overview	Popularity	Vote_Count	Vote_Average	Original_Language	Genre
25545	2021-03-31	The United States vs. Billie Holiday	Billie Holiday spent much of her career being ...	13.354	152	(6.5, 7.1]	en	Music
25546	2021-03-31	The United States vs. Billie Holiday	Billie Holiday spent much of her career being ...	13.354	152	(6.5, 7.1]	en	Drama
25547	2021-03-31	The United States vs. Billie Holiday	Billie Holiday spent much of her career being ...	13.354	152	(6.5, 7.1]	en	History
25548	1984-09-23	Threads	Documentary style account of a nuclear holocau...	13.354	186	(7.1, 10.0]	en	War
25549	1984-09-23	Threads	Documentary style account of a nuclear holocau...	13.354	186	(7.1, 10.0]	en	Drama
25550	1984-09-23	Threads	Documentary style account of a nuclear	13.354	186	(7.1, 10.0]	en	Science Fiction

Which year has the most filmed movies?

```
In [34]: df['Release_Date'].hist()  
plt.title('release date col distribution')  
plt.show()
```



Conclusion

Q1: What is the most frequent genre in the dataset?

- Drama genre is the most frequent genre in our dataset and has appeared more than 14% of the times among 19 other genres.

Q2: What genres has highest votes?

- we have 25.5% of our dataset with popular vote (6520 rows). Drama again gets the highest popularity among fans by being having more than 18.5% of movies popularities.

Q3: What movie got the highest popularity? what's its genre?

- Spider-Man: No Way Home has the highest popularity rate in our dataset and it has genres of Action, Adventure and Sience Fiction.

Q4: What movie got the lowest popularity? what's its genre?

- The united states, thread' has the highest lowest rate in our dataset and it has genres of music, drama, 'war', 'sci-fi' and history".

Q5: Which year has the most filmed movies?

- year 2020 has the highest filming rate in our dataset.