Netflix Movie Data Analysis Project

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.

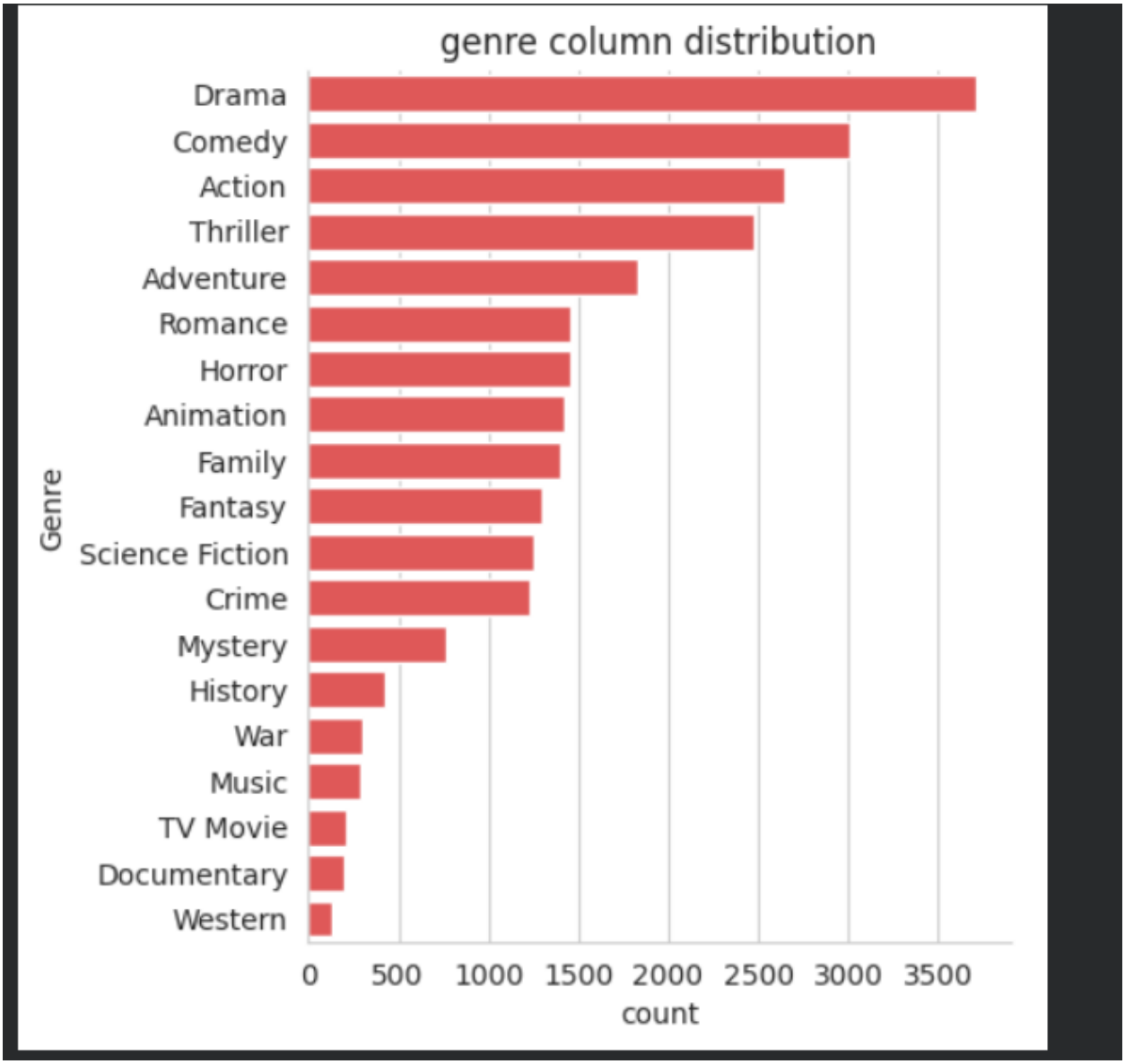
1. What is the most frequent genre of movies released on Netflix?
2. Which has highest votes in vote avg column?
3. What movie got the highest popularity? what's its genre?
4. What movie got the lowest popularity? what's its genre?
5. Which year has the most filmed movies?

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

```
# showing stats. on genre column
df['Genre'].describe()
```

	Genre
count	25552
unique	19
top	Drama
freq	3715

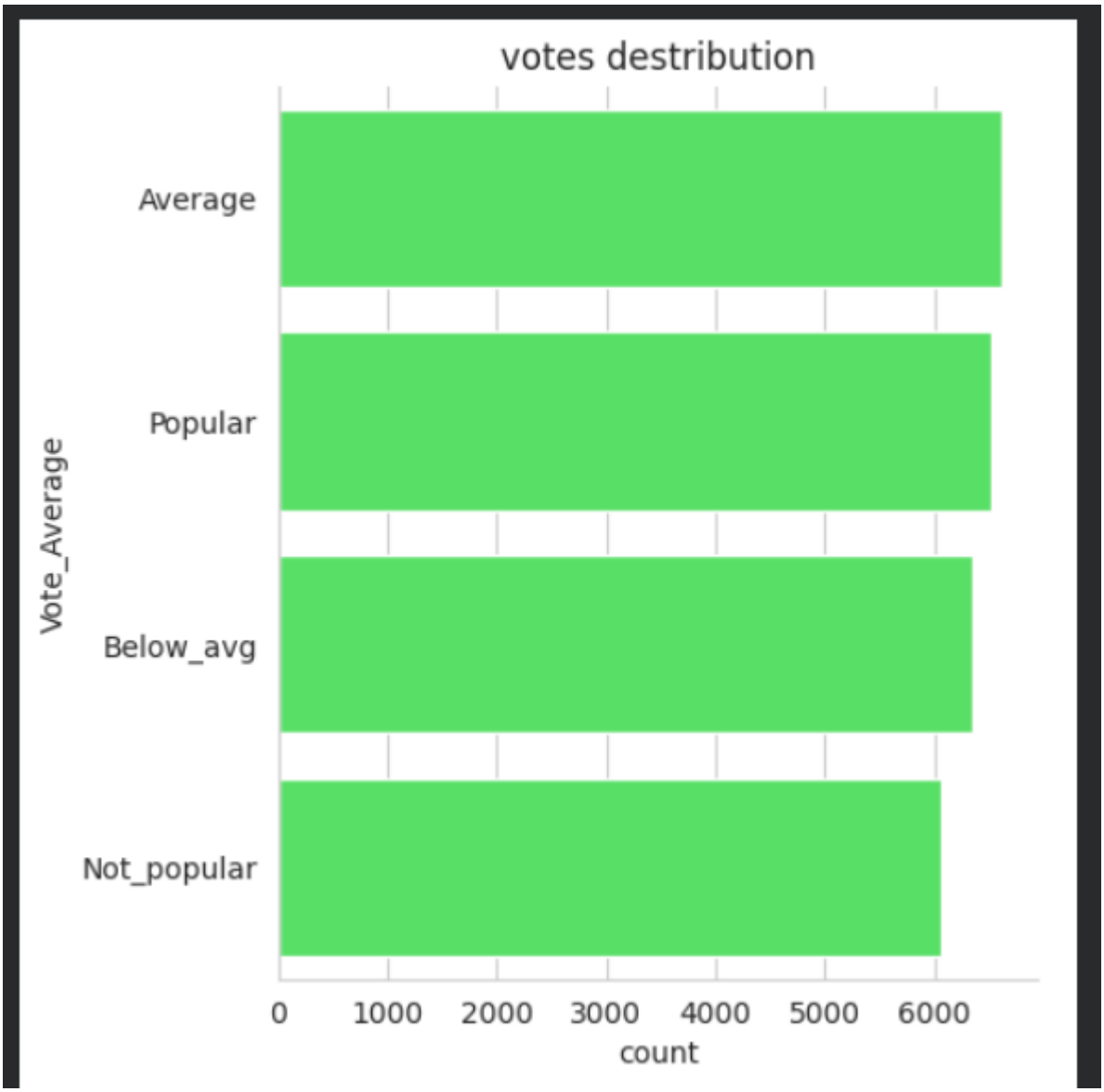
```
sns.catplot(
    y='Genre',                  # categorical variable on Y-axis
    data=df,                    # DataFrame source
    kind='count',               # count plot (frequency of each category)
    order=df['Genre'].value_counts().index,  # order by frequency
    color='#f54242')            # bar color
plt.title('genre column distribution')
plt.show()
```



Conclusion: 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?

```
# visualizing vote_average column
sns.catplot(y = 'Vote_Average', data = df, kind = 'count',
order = df['Vote_Average'].value_counts().index,
color = '#42f554')
plt.title('votes destribution')
plt.show()
```



Output: 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 it's genre?

```
# checking max popularity in dataset
df[df['Popularity'] == df['Popularity'].max()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre
0	2021	Spider-Man: No Way Home	5083.954	8940	Popular	Action
1	2021	Spider-Man: No Way Home	5083.954	8940	Popular	Adventure
2	2021	Spider-Man: No Way Home	5083.954	8940	Popular	Science Fiction

Output: Spider-Man: No Way Home has the highest popularity rate in our dataset and it has genres of Action, Adventure and Science fiction**.**

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

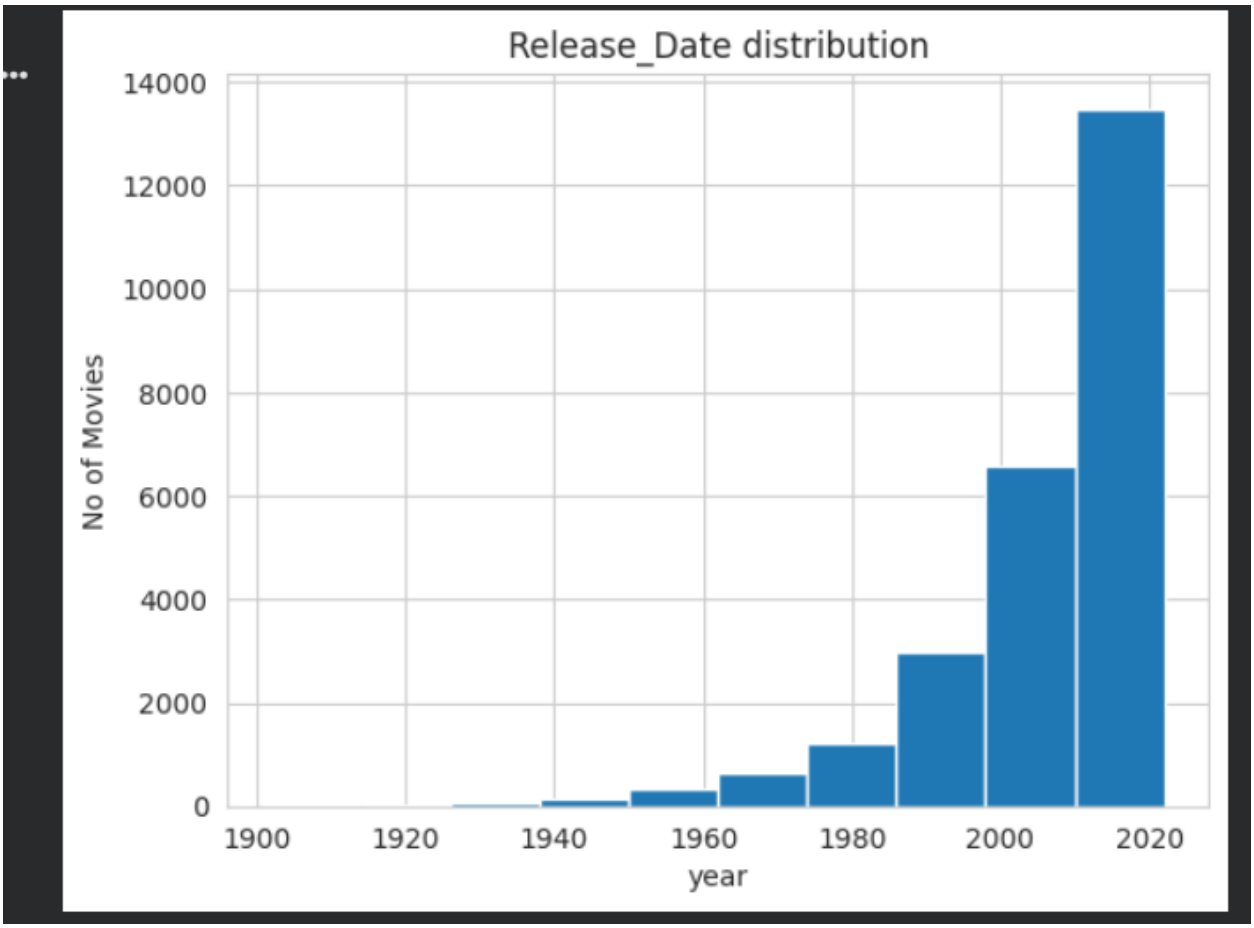
```
# checking min popularity in dataset
df[df['Popularity']==df['Popularity'].min()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre
25546	2021	The United States vs. Billie Holiday	13.354	152	Average	Music
25547	2021	The United States vs. Billie Holiday	13.354	152	Average	Drama
25548	2021	The United States vs. Billie Holiday	13.354	152	Average	History
25549	1984	Threads	13.354	186	Popular	War
25550	1984	Threads	13.354	186	Popular	Drama
25551	1984	Threads	13.354	186	Popular	Science Fiction

Output: 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?

```
df['Release_Date'].hist()
plt.title('Release_Date distribution')
plt.ylabel('No of Movies')
plt.xlabel('year')
plt.show()
```



Output: year 2020 has the highest filming rate in our dataset**.**