

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

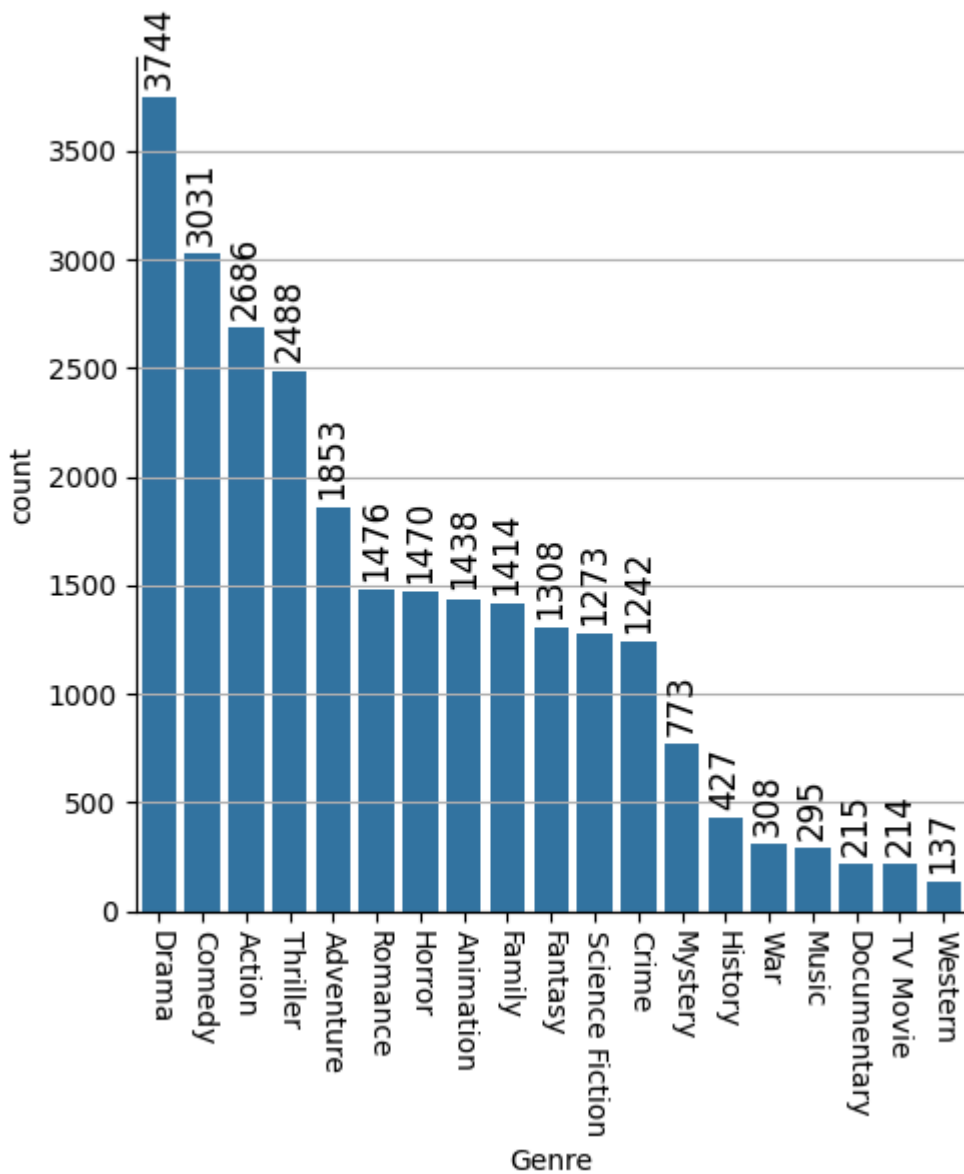
```
In [27]: df.loc[:, "Genre"].nunique()
```

```
Out[27]: 19
```

```
In [28]: g = sns.catplot(x= "Genre",kind="count",data=df,order=df['Genre'].value_counts().index)
plt.grid(axis="y")
plt.xticks(rotation=270)

# Add Labels on top of bars
ax = g.ax # Get the axes from the FacetGrid
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f", label_type="edge", fontsize=12, color='black', padding=3, rotation=90)

plt.show()
```



- we can notice from the above visual that 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 ?

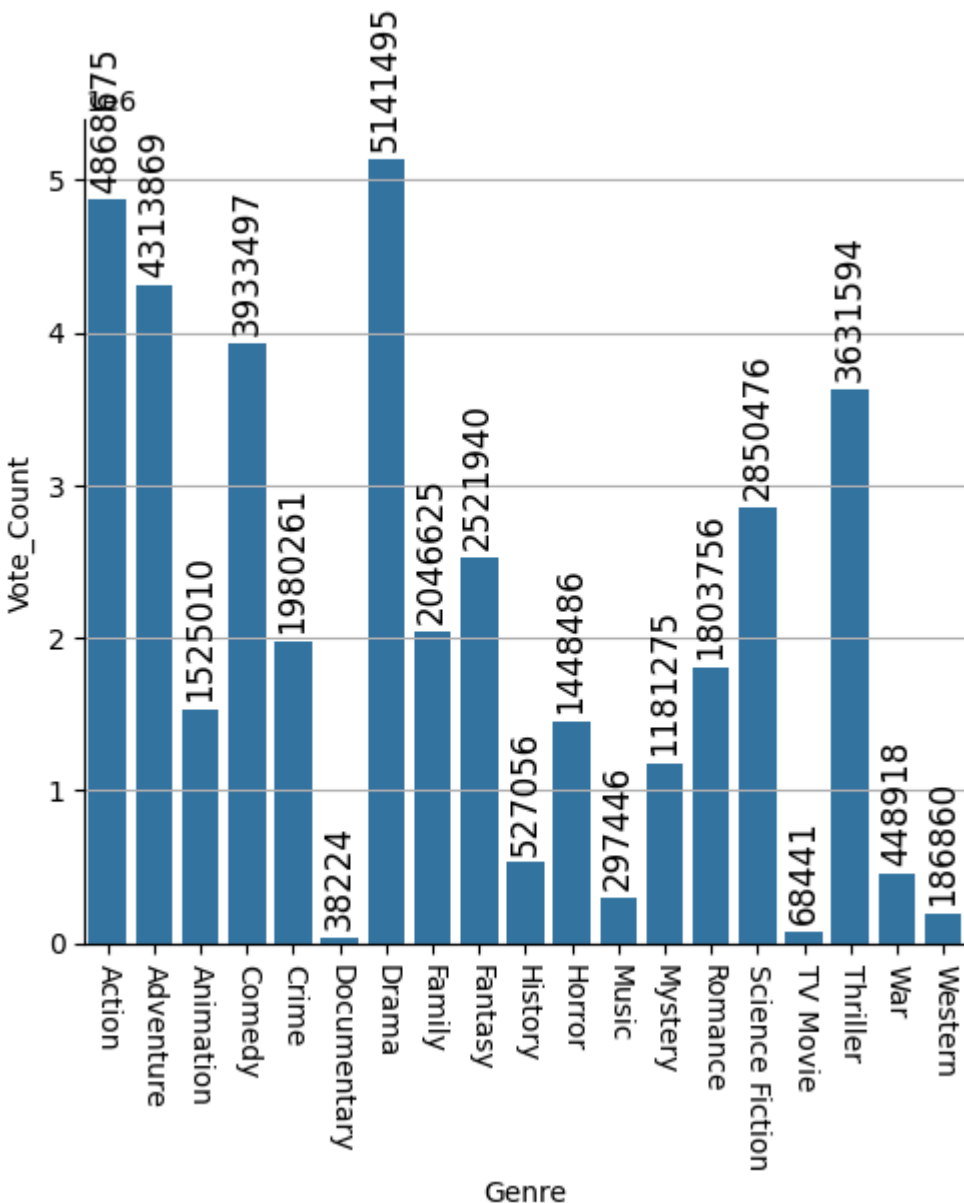
```
In [29]: df.columns
```

```
Out[29]: Index(['Release_Date', 'Title', 'Popularity', 'Vote_Count', 'Vote_Average',
              'Genre', 'Day', 'Month', 'Year'],
              dtype='object')
```

```
In [30]: g = sns.catplot(x= "Genre",y="Vote_Count",kind="bar",data=df,errorbar=None,estimator=np.sum)
plt.grid(axis="y")
plt.xticks(rotation=270)

# Add Labels on top of bars
ax = g.ax # Get the axes from the FacetGrid
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f", label_type="edge", fontsize=12, color='black', padding=3, rotation=90)

plt.show()
```



- Drama Genre has the most vote count

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

```
In [31]: df.columns
```

```
Out[31]: Index(['Release_Date', 'Title', 'Popularity', 'Vote_Count', 'Vote_Average',
              'Genre', 'Day', 'Month', 'Year'],
              dtype='object')
```

```
In [34]: df.loc[:, "Popularity"].dtype
```

```
Out[34]: dtype('float64')
```

```
In [33]: df.loc[:, "Popularity"].max()
```

```
Out[33]: 5083.954
```

```
In [38]: df[df.loc[:, "Popularity"] == df.loc[:, "Popularity"].max()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre	Day	Month	Year
0	2021-12-15	Spider-Man: No Way Home	5083.954	8940	8.3	Action	15	12	2021
1	2021-12-15	Spider-Man: No Way Home	5083.954	8940	8.3	Adventure	15	12	2021
2	2021-12-15	Spider-Man: No Way Home	5083.954	8940	8.3	Science Fiction	15	12	2021

```
In [40]: df[df.loc[:, "Popularity"] == df.loc[:, "Popularity"].max()][["Title", "Genre"]]
```

	Title	Genre
0	Spider-Man: No Way Home	Action
1	Spider-Man: No Way Home	Adventure
2	Spider-Man: No Way Home	Science Fiction

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

- Drama Genre has the most vote count

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

```
In [42]: df.loc[:, "Popularity"].min()
```

```
Out[42]: 13.354
```

```
In [44]: df[df.loc[:, "Popularity"] == df.loc[:, "Popularity"].min()]
```

	Release_Date	Title	Popularity	Vote_Count	Vote_Average	Genre	Day	Month	Year
25786	2021-03-31	The United States vs. Billie Holiday	13.354	152	6.7	Music	31	3	2021
25787	2021-03-31	The United States vs. Billie Holiday	13.354	152	6.7	Drama	31	3	2021
25788	2021-03-31	The United States vs. Billie Holiday	13.354	152	6.7	History	31	3	2021
25789	1984-09-23	Threads	13.354	186	7.8	War	23	9	1984
25790	1984-09-23	Threads	13.354	186	7.8	Drama	23	9	1984
25791	1984-09-23	Threads	13.354	186	7.8	Science Fiction	23	9	1984

```
In [41]: df[df.loc[:, "Popularity"] == df.loc[:, "Popularity"].min()][["Title", "Genre"]]
```

	Title	Genre
25786	The United States vs. Billie Holiday	Music
25787	The United States vs. Billie Holiday	Drama
25788	The United States vs. Billie Holiday	History
25789	Threads	War
25790	Threads	Drama
25791	Threads	Science Fiction

- The United States vs. Billie Holiday and threads these 2 movies have lowest popularity
- Movie "The United States" has Genre of Music, Drama and History
- Movie "The United States" has Genre of War, Drama and Science Fiction

Q5: Which year has the most filmed movies?

```
In [45]: df.columns
```

```
Out[45]: Index(['Release_Date', 'Title', 'Popularity', 'Vote_Count', 'Vote_Average',
              'Genre', 'Day', 'Month', 'Year'],
              dtype='object')
```

```
In [53]: np.round(df.loc[:, "Year"].value_counts() / df.loc[:, "Year"].value_counts().sum(), 5) * 100
```

```
Out[53]: Year
2021    6.351
2018    5.378
2017    5.292
2019    4.932
2016    4.699
...
2024    0.016
1902    0.012
1925    0.008
1929    0.008
1930    0.008
Name: count, Length: 102, dtype: float64
```

- In year 2021 6.35% of movies out of total were released highest ever in a year

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

END: